MULTI-LABEL INCREMENTAL KERNEL EXTREME LEARNING MACHINE FOR FOOD RECOGNITION

CHEN SAI

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYA
KUALA LUMPUR

2022

MULTI-LABEL INCREMENTAL KERNEL EXTREME LEARNING MACHINE FOR FOOD RECOGNITION

CHEN SAI

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE

**FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYA
KUALA LUMPUR**

2022

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: CHEN SAI

Registration/Matric No: WOA170021

Name of Degree: Master of Computer Science (Applied Computing)

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

MULTI-LABEL INCREMENTAL KERNEL EXTREME LEARNING MACHINE FOR FOOD RECOGNITION

Field of Study: MACHINE LEARNING

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date 2022.1.28

Subscribed and solemnly declared before,

Witness's Signature

Date

Name:
Designation:

[MULTI-LABEL INCREMENTAL KERNEL EXTREME LEARNING MACHINE FOR FOOD RECOGNITION]

ABSTRACT

Real-world food datasets are not fixed, it is open-ended and dynamic, however, the novel machine learning methods for food recognition have poor performance in incremental learning datasets. If food samples and food categories continuous increase, these methods may need to train again from the beginning. This is time-consuming and occupies computational resources. My study proposed a multilabel classifier for this shortcoming, called Multi-Label Adaptive Reduced Class Incremental Kernel Extreme Learning Machine, the abbreviation is ARCIKELM-ML. We applied Inception-Resnet-V2 for food feature extraction and the Relief F method for feature ranking and selection. Then used ARCIKELM-ML for multi-label classification. In the framework, the hidden and output neurons corresponding to new labels are added and the classifier progressively remodels its structure like the new labels are introduced from the beginning of the training process. The experiment for food ingredients recognition is based on three standard benchmark datasets and evaluated on F1 score, Hamming Loss, Recall Score and Precision Score. Results showed that the proposed ARCIKELM-ML algorithm has good performance and meets the criteria of incremental learning

Keywords: Kernel Extreme Learning Machine, Incremental Learning, Multi-label classification, Food Ingredients Recognition

[MESIN PEMBELAJARAN EKSTREME KERNEL MULTILABEL INCREMENTAL UNTUK PENGIKTIRAFAN MAKANAN]

ABSTRAK

Set data makanan di dunia nyata tidak tetap, bersifat terbuka dan dinamik, namun kaedah pembelajaran mesin baru untuk pengkelasan makanan mempunyai prestasi yang rendah dalam set data pembelajaran tambahan. Sekiranya sampel makanan dan kategori makanan terus meningkat, kaedah pembelajaran mesin mungkin perlu dilatih dari permulaan semula. Ini memakan masa dan menggunakan sumber komputasi. Dalam projek ini saya mencadangkan pengkelasan multilabel untuk kekurangan ini, yang disebut "Multi-Label Adaptive Reduced Class Incremental Kernel Extreme Learning Machine" (ARCIKELM-ML). Kami menggunakan Inception-Resnet-V2 untuk feature extraction makanan dan kaedah ReliefF untuk ranking dan feature selection. Seterusnya menggunakan ARCIKELM-ML untuk klasifikasi pelbagai label. Dalam kerangka kerja, hidden neuron dan output yang sesuai dengan label baru boleh ditambah secara berterusan dan pengklasifikasi secara progresif mengubahsuai strukturnya sekiranya label makanan baru diperkenalkan semasa proses latihan. Eksperimen untuk pengkelasan ingredient makanan dibuat berdasarkan tiga set data penanda aras standard dan dinilai pada skor F1, Hamming Loss, Recall Score dan Precision Score. Hasil kajian menunjukkan bahawa algoritma ARCIKELM-ML yang dicadangkan mempunyai prestasi yang baik dan memenuhi kriteria pembelajaran secara berterusan. Keywords: Mesin Pembelajaran Ekstrem Kernel, Pembelajaran Tambahan, Pengelasan pelbagai label, Pengkelasan Bahan Makanan

Keywords: Mesin Pembelajaran Ekstrem Kernel, Pembelajaran Tambahan, Pengelasan pelbagai label, Pengkelasan Bahan Makanan

ACKNOWLEDGEMENTS

Time flies, when looking back on the time in the past few years, I feel very rewarding. Get in touch with the most cutting-edge scientific research and knowledge, such as machine learning, computer vision and natural language processing and so on. At the same time, it has also enriched my life experience. During the campus life, I called for the establishment of the Chinese Students and Scholars Association University of Malaya and served as the first Chinese Student Chairperson.

I would like to thank my parents first for providing the material foundation and spiritual encouragement, so that when I study in a foreign country, I do not need to worry about food and clothing and I can study with peace of mind. Secondly, I would like to thank my supervisor, Professor Dr Loo Chiukiong, who led me into the field of artificial intelligence research, which laid the foundation for my future research, cultivated independent scientific research methods and a strong interest in the exploration of new knowledge. At the same time, I also want to thank the seniors in our laboratory, Mr Tahir Ghalib Ahmed and Dr Liu Zongying, who guided me and help in my study and life in the laboratory life. Many friends have enriched me abroad life and we have experienced foreign culture and food together. I could not thank them one by one here. At the end of my study abroad career, I would like to say thank you sincerely to all of them!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

(K)ELM	:	(Kernel) Extreme Learning Machine
F(x)	:	Mapping function
FP	:	False Positive
FN	:	False Negative
H	:	Non-singular matrix
H^+	:	Moore-Penrose generalized inverse of H
$k_b; k_n$	:	Initial kernel and undated kernel matrix
L	:	Number of hidden layer nodes
M	:	Transformation Matrix
N	:	Neuron eligibility Matrix
$\mathbb{N}, \mathbb{N}_b$	:	Labels set and selected labels
O_j	:	learning error
TN	:	True Negative
TP	:	True Positive
$X_b; X_n$	:	The samples of based labels and new labels
Z_n	:	Newly added hidden nodes of output matrix
$\beta^{(b)}\beta^{(n)}$	:	Output weights of Initialization and updated output weights
μ_{ih}	:	Membership Value

CHAPTER 1: GENERAL INTRODUCTION

1.1 Overview

This chapter is to give an overview of the research background, a general introduction of the relevant area, problems setting, what this thesis aims for and a thesis outline. The study first gives a general overview of my research background about extreme learning machines, multi-label classification and food recognition. Then the study states the research problems, research objectives separately. The proposed methods and research contributions are also introduced. The final part of this chapter is the introduction of the organizations of the dissertation.

1.2 Extreme Learning Machine

In recent decades, the field of artificial intelligence has developed rapidly, and machine learning has been applied to a wide range of fields, scenarios and applications, such as the commodity recommendation system for online shopping, the electronic payment system for face recognition, the task-based dialogue system. Machine learning greatly improves business decisions and productivity.

In the machine learning algorithm, the extreme learning machine (ELM) algorithm has been paid much attention by plenty of scholars. ELM algorithm is regarded as single hidden layer feedforward neural network (SLFNs) learning algorithm which has outstanding performance. The researchers can analytically determine the parameters of ELMs instead of tuning them, so it tends to learn extremely fast and have better generalization performance. In recent years, many scholars have optimized and improved ELMs, such as optimised the kernels and implemented the incremental learning on the ELMs. Researchers have applied ELMs to classification, regression, feature extraction, clustering and achieved good results. In the following chapters, the study will mainly discuss the application of ELM in classification research.

1.3 Multi-Label Classification

Classification tasks have a wide range of applications in machine learning. In most traditional classification tasks, each sample is only associated with one or two categories or labels. Researchers have proposed many efficient and accurate algorithms, such as Naive Bayes, Support Vector Machine (SVM) and decision tree methods. However, data has a variety of semantic information, and these methods cannot accurately describe the actual object. For example, a piece of news can be labelled with several labels at the same time, such as "education", "finance", "policy", "University of Malaya".

Different from single-label classification and binary classification, instances in multi-label classification are usually associated with multiple disjoint labels. Therefore, the multi-label classification task is more in line with the laws and characteristics of the objective world. At the same time, multi-label classification is ubiquitous in real scenarios, such as text classification, sentiment classification, music categorization, and semantic scene classification. The study also had further discussion this in the literature review section.

1.4 Food Recognition

People's demands for food have not just eliminated hunger, using modern technology to improve the living standard and the quality of diet have been the targets. In addition, the researchers also need to pay more attention to food composition for some patients such as people with hypertension, heart attack, high cholesterol and diabetes.

Research on the combination of food-related research and computer science has become more and more popular in recent years. Smart devices that combine food-related research with computer vision, machine learning, and social networks are also emerging in an endless stream. The large amounts of data generated and collected by these smart devices have also promoted the development of many food-related fields, such as food science, gastronomy, agriculture, food security and food computing, etc. In recent years, a large number of scholars have introduced deep learning and reinforcement learning methods to the field of image recognition. They have reduced the cost of training sample labelling through transfer learning, few-shot learning and other methods. At the same time, it has brought about an increase in training speed and accuracy. Among food-related tasks, common tasks include food-oriented perception, food recognition, recipe retrieval, food-related recommendation and so on. The research is mainly about food ingredients recognition and classification.

In health systems, developers used artificial intelligence as foundation support for food applications. To give a food picture, the systems can automatically recognize food type, quantity, estimate calories and nutrients to help users manage their diet. It can also predict food ingredients by multi-label classification and recommend similar recipes to them. Therefore, the research on food recognition technology has extremely important research value and practical significance, and a food recognition system also has a strong market prospect.

1.5 Problem Statements

In recent years, there are many articles about food recognition, but seldom about incremental food ingredients classification and the previous classification methods are slow (Wang, Y. et al. 2019). Such classifiers work well with the pre-known datasets, but they may not be appropriate for applications such as food recognition with the attributes of training data is not clear (Venkatesan, R. et al. 2016). Moreover, the data is collected and it will not be added up, but in reality, the data for many projects is constantly increasing. Training with the existing data is insufficient to predict and solve real problems. Nevertheless, the past food recognition work was mostly carried out on fixed data sets, which have a high diversity in food categories at the start.

The researchers assumed their training datasets have all the food classes or limited them to small categories. There is a pressing need to organize it because, in the real-world setting, the available training data is continuously increasing. Besides, modern classification tasks usually predict multiple labels associated with an instance at the same time (Wang, Y. et al. 2019). The new concepts occur over time and different people have different concepts of the same thing. However, food ingredients classification can be catered by using multi-label incremental learning in real-world scenarios. Some researchers use multi-class incremental methods for food recognition, however, the algorithms in their paper (Tahir, G. A. et al. 2020) cannot recognize food ingredients and cannot add new labels corresponding to incoming ingredients during incremental learning.

Therefore, the research problems are as below:

a) Is that possible to propose an incremental network for the food ingredients multi-label classification based on the kernel extreme learning machine?
b) How to dynamically deal with the multi-label features when the food ingredient labels are continuously increasing in the whole network?

1.6 Research Objectives

According to the research problem statements above, the previous ingredients classification methods are slow, the study will use KELM to replace the traditional feedforward neural networks in the training process. This will decrease training time and it has already been confirmed by previous paper (Huang G.B. et al. 2006).

The study is going to identify food ingredients presented on one plate by using a multilabel classifier based on the ARCIKELM. In the real scenario, the data usually increase and the attributes of new data are unknown. In order to adopt the methods and classifier to this situation. Therefore, the food ingredients classification system has to satisfy the criteria of incremental learning. Further More, the classifier should progressively remodel its structure like the new labels are introduced from the beginning of the training process. In addition, it should be able to learn newly added labels incrementally in real-time. At the same time, considering data security and model complexity, the system does not need to store the images.

To summary up, the research objectives are as following:

a) To propose an incremental network for the food ingredients multi-label classification based on kernel extreme learning machine.
b) To compare and evaluate the classifiers on standard benchmark datasets and visualize the increasing labels.

1.7 Proposed Methods

The study proposed a novel multilabel classifier using Adaptive Reduced Class Incremental Kernel Extreme Learning Machine (ARCIKELM-ML) for food ingredients classification. This classification method can meet the requirements of incremental multi-label classification problems in the real world, and at the same time, it has good food ingredients recognition and classification capabilities for the newly added labels.

The study first implements Inception-Resnet-V2 for food feature extraction and the Relief F method to reduce the dimensions and complexity of the features by ranking and selecting the best representations. After the pre-train and fine-tune model, the features were implemented on the multi-label classification, the new labels are gained when they are added. For further explanation, the output neurons corresponding to new labels are added and the classifier can progressively remodel its structure like the new labels are introduced from the beginning of the training process. The study proves the effectiveness of the method through experiments on three standard datasets and visualises the results on food images.

1.8 Research Contributions

There are a lot of studies and applications for food recognition, but the algorithms are all based on a certain dataset and cannot adapt to new data that all data needs to be retrained. Therefore, an algorithm that can adapt to dynamic data is urgently needed. The algorithms in the study just make up for this shortcoming.

The study tries to combine Extreme Learning Machine and incremental learning algorithm. It provides a multi-label classification method for incremental research in the field of food ingredients recognition. To our knowledge, there was no research uses incremental ELM to do food ingredients multi-label classification. The proposed method ARCIKELM-ML could adjust the structure dynamically when new labels were added. This could be used more in food energy analysis, nutritional analysis, food testing and user diet management in the future. At the same time, it also provides good inspiration for multi-classification algorithms and provides a reference for researchers in dynamic multi-classification real-world problems.

In the study, the author's contributions were to design the downstream work of the networks and wrote the multi-label classification codes on Google Colab. The pre-process of the images and some visualization work was done in the training process. Besides, set up the experiments and analyse the multilabel classification results. The features extraction and selection were based on previous studies which were cited in the following parts.

1.9 Thesis Organization

The dissertation is divided into five chapters and there is an overview in each section. The structure is as follows.

Chapter one is a general introduction. To give an overview of the research background, a general introduction of ELM, multilabel classification and food recognition, the research problems and objectives, proposed methods and research contributions.

Chapter two is the literature review. To give details of the basic theory, research methods and research histories of the ELM, multilabel classification and food recognition. At the same time, some studies on the combination of the three are also discussed.

Chapter three is methodology. To introduce the system design and the three benchmark datasets, Vireo Food Dataset, Food 101 and Simplified Food101. There are three proposed methods, which are Inception ResnetV2, Relief F and Adaptive Reduced Class Incremental Kernel Extreme Learning Machine for multi-label classification. Meanwhile, the reasons for choosing these methods are also expounded.

Chapter four is about the experiment design and experiment environments. The experimental evaluation criteria are also introduced. The study chose Precision Score, Recall Score, F1 Score and Hamming Loss as the measurement methods. In addition, the comparison and analysis of the classification results are discussed in this chapter.

In the end, chapter five summarizes the whole dissertation and points out the problems and directions of future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

In recent years, with the development of artificial intelligence and the gradual maturity of image processing technology, scholars have done extensive and in-depth research on food recognition. This research is about multi-label food ingredients classification based on the incremental kernel extreme learning machine.

The kernel extreme learning machine is the basis for the research on multi-classification and incremental algorithms. In the Section 2.2, the study first introduces the researches of food ingredients recognition and classifications based on the kernel extreme learning machine. Then explain the principle of the kernel extreme learning machine furtherly, as well as its variants such as the incremental kernel extreme learning machine. Finally, the basic principles and applications of multi-classification algorithms in recent years are discussed.

2.2 Food Ingredients Recognition and Classification

In the study of food ingredients classification, researchers usually put food recognition and classification together, because food recognition will affect the results of food ingredients classification. Food recognition is the basis and premise of food classification. In order to clarify the structure of the article, the study will discuss Food Ingredients Recognition and classification separately. In the food recognition part, the study also discusses popular food recognition methods, recent research hotspots and applications. The food classification part is mainly about common classification methods and multi-label classification. Finally, the study points out the shortcomings of the existing algorithms, leading to the incremental learning classification algorithm.

2.2.1 Food and Ingredients Recognition

Nowadays, people have higher requirements for food intake. Food recognition is a favourite topic in computer vision that has received attention from academia and industry. Food recognition and composition analysis can provide people with reliable dietary recommendations. People develop health-oriented systems, such as an automated dietary planar app, to improve people's dietary habits. The users can take a food picture or insert their daily food intake, then the app can automatically recognize (Meyers et al. 2015) the food type and quantity, retrieve relevant images and recipes, and also estimate consequent calories and nutrition intake. (Battiato, S. et al. 2021) extracted information from food picture and contextual data to analyze the correlations between eating habits and smoke quitting protocols.

Most recently, (Manjunathan, A. et al. 2021) proposed a capable Fast Deep Learning method to calculate the nutrition and calorie from food intake of patients and dietitians every day. In recent years, image processing technology has gradually matured, and the accuracy of image recognition of many models exceeds the level of humans. However, the depth of the network framework has also greatly increased, which inevitably leads to huge consumption of computing resources. (He, K. et al., 2015) applied deep residual learning framework to ease the training of networks. ResNet-50 means 50-layer residual network which used many bottleneck blocks are stacked together. The study also implemented to extract image features in the experiments. Instead of using multiple stacked layers to directly fit the desired feature map, the study explicitly used them to fit a residual map. One of the characteristics of the residual network is the jump connection, which can make the information of the previous residual block flow into the next residual block without hindrance, improve the flow of information, and also avoid the vanishing gradient caused by the network being too deep. Problems and degradation problems. (Huang, G. et al. 2017) introduced Dense Convolutional Network (DenseNet), where layers are directly linked through feed-forward, and the feature maps of all previous layers are used as input. This network structure has many advantages, it remitted the vanishing-gradient problem, strengthened feature propagation, encouraged feature reuse, and required fewer parameters and computation. (Szegedy, C. et al., 2017) applied inception architecture with residual connections which accelerated the training of inception networks significantly. The Inception-ResNet-v2 is a costlier hybrid Inception version with significantly improved recognition performance. (Wang, J. et al. 2016) proposed a unified framework that combined recurrent neural networks (RNNs) and convolutional neural networks (CNNs). They used this CNN-RNN network to learn semantic label dependencies and relevance from scratch in an end-to-end fashion. Results on standard datasets showed that this method had better performance in multi-label classification work.

In recent years, many scholars investigate in food-related surveys. (Salim, N. O. et al. 2021) compared a lot of recent articles on the deep learning methods for the image processing of food pictures, text and spectrum. The authors also proposed an approach for automating data collection and build an automatic analysis pipeline to identify the interest and predict the food classes. (Harper et al. 2015) proposed a food computing survey which is mainly in the view of the Open Agriculture field. They wanted to build the next generation distributed farming system using sensing and data collection. (Knez, S. et al 2015) studied mobile food recognition systems and applied nine recent research on detecting eating activity based on their architecture. They used nine application as monitors to record the food items and stored online. Users can get the nutrition information of selected food. This system is costly and not accuracy because they cannot estimate volume of different shape and size. (Trattner, C. et al. 2017) summarized the novel food recommender systems and resources for researchers and highlight important and recent approaches to food recommendation. They also examined the algorithms have been implemented in the food area and pointed out the future directions. As we know, if people keep a healthy diet and balanced nutrition, they will probably stay away from disease, (Bruno et al. 2017) gave a survey on automatic food monitoring and diet management system using wearable devices and applications in smartphone. The researchers presented a lot of methodologies and resources on this topic. These surveys are limited to a small area, and there is no comprehensive and in-depth analysis of food-related research. (Min, W. et al. 2019) also proposed a survey on food computing. The difference is that they combined food computing and computer science to the comprehensively and in-depth analysis of current efforts in food computing. In total, they studied nearly 300 studies, summarized food-related datasets, tasks and applications, and discussed the key challenges of food identification.

In the real world, there are many types of food and people often need advice when choosing what to eat. Many factors affect the choice, such as personal tastes, people's comment, nutritional information, food colour and fragrance, etc. If the application can provide this information, this will help people estimate their food calories for keeping fit and analyse their eating habits for healthcare. There are already many recipes recommend systems through seeking for the best set of ingredients combined with some comprehensive information to assist users in discovering their favourite foods, and also provide health dietary recommendations for some sick users.

In recent years, there were a lot of researchers investigated the food-related research area. Some scholars have done in-depth research on food category recognition. With an intuition of network architecture design should meet the attributes of food composition (Martinel, N. et al. 2018) leveraged deep Wide-Slice Residual Network to handle the food structure and combine with the sliced convolution block to produce the classification score. Inspired by the residual deep network, the author applied a slice convolution block to handle the food structure. The outputs of the network are the prediction scores for the food categories. Their experimental results show that their methods have better performance than normal hand-crafted features or deep learning schemes. However, their methods cost much computation resources and cannot be implemented on mobile platforms. (Chen, M. Y. et al. 2012) combined multi-label SVM classifier with multi-class Adaboost algorithm for automatic food categories identification using feature descriptors, such as SIFT, colour and texture feature descriptor. Besides, they also estimated quantity based on depth information. (Bossard, L et al. 2014) applied Random Forests for automatically recognizing domain food categories in food dishes. They identified discriminative image regions to easily distinguish the dish type. Meanwhile, they only aligned patches with image super pixels to improve the speed of mining and classification. Differs from this kind of work, (Chen J. et al. 2016) proposed architecture for joint relationship of food categorization and food ingredients recognition, they combined semantic labels of ingredients with deep food image feature and applied them for zero-shot retravel. Their model produced an impressive performance on food recipe recognition, however, could not deal with unseen ingredients labels.

There have been a lot of technical methods to realize the application of food recognition and food composition recognition. The deep learning approach utilizes the ability to learn features from label data to achieve the most advanced performance in food recognition. For this matter, the food recognition algorithm can be divided into two categories, one is based on network structure methods, another is based on visual features methods.

In the network structure methods, (Bolaos M et al. 2017) proposed a combination of multiple classifiers based on the convolution model, which complemented each other with the depth of the model to improve performance. Compared with traditional image recognition technology, food recognition is more difficult. (Kawano, Y. et al. 2014) combined Deep Convolutional Neural Network with hand-crafted features and Fisher Vectors to boost the accuracy of extracting food image features. Besides, (Kawano, Y. et al. 2013) proposed real-time food item recognition applications on a smartphone that needs the user to draw bounding boxes. In the same way, (Girshick, R. et al. 2014) proposed a very effective classifier that combined Regions with CNN features called Faster R-CNN for multiple food recognition, this method also detecting objects with its bounding boxes. (Kawano Y. et al. 2014) fine-tuned and pre-trained Deep Convolutional Neural Network (DCNN) and made an optimal combination of the activation characteristics extracted from the pre-trained DCNN. They found that DCNN is suitable for large-scale data and still have high accuracy in classification.

In the visual feature-based methods, (Tatsuma A. et al. 2016) put forward a new image representation method composed of the covariance of Convolutional Layer feature maps, which exceeded the previous bag-of-Visual-words histogram, improved Fisher vector, CNN-SVM and other methods. CNN showed that it was significantly higher than the traditional SVM-based manual functional method. They also found that colour plays a leading role in the feature extraction process of convolution kernel, and the accuracy of CNN in food image detection is also significantly higher than that of traditional methods.

The study mainly discussed feature extraction and feature selection algorithms which commonly used in image processing. The study did not use the ingredients features directly but chose the best features first. This is also a very important part of the experiment. The study also took many preprocess methods to get the best images features before food recognition. The study mainly discussed the related research of Inception-Resnet-V2 and Relief F. (Szegedy, C. et al. 2017) compared several variants of Inception and Inception-ResNet deep networks. The Inception-ResNet-v2 and Inception- v4 have comparable results in the validation dataset of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). However, the speed of training Inception-v4 is slower and the framework is more complex. The study used Inception-ResNet-v2 as feature extraction model, which has a significantly high recognition score. Further, Inception-ResNet-V2 accelerates the training process which combined the benefits from both ResNet and Inception Models.

Among the feature selection methods of interaction with the learning algorithm, methods such as a wrapper, embedded and filter are more common. Wrappers usually evaluate feature importance using some specific learning algorithms. Unlike the wrapper method, the embedded approach combines feature selection with the training process of a learning algorithm. Unlike the above two approaches, filters usually use feature ranking to remove irrelevant or redundant features no matter how is the learning algorithm. Some algorithms have good results, such as the Binary Relevance (BR) approach and Label Powerset (LP) approach, but they are mainly for single-label learning, the study will not discusses them in-depth here. (Spolaôr, N. et al.2013) proposed Relief F method directly for multi-label feature selection, experimental results in synthetic datasets show that RF-ML ranks the relevant features as the best ones more often when compared to the single-label Relief F algorithm in which the multi-label data has been transformed using two well-known data transformation approaches. For Relief F, F means the sixth version of the Relief-based algorithm, which is widely used in estimating the quality of features in problems. It is efficient to detect the conditional dependencies between features and robust to deal with incomplete and noisy data. Relief F can combine the contextual information to do feature ranking and selection. (Robnik-Šikonja, M. et al. 2003) investigated on the survey of Relief F at the theoretical and empirical level. They described and introduced the principles and details of the Relief F algorithm in a comprehensive and in-depth manner. (Urbanowicz, R. J. et al. 2018) is the latest survey on feature selection using Relief-based approaches, which compared the novel Relief family algorithms in many aspects, such as contributions, strategies, functionality, time complexity, adaptation to key data characteristics, and software availability.

When make food recognition, a good food dataset is usually very important, which may directly affect the quality of the experimental results. Scholars created different public datasets according to different research objectives and research fields. (Bossard, L et al. 2014) released the FOOD101 dataset, and the experiment also based on this dataset. The specific situation can be seen in the experimental part. (Xu, R. et al. 2016) published a dataset of restaurants with geolocation and external information. Based on this dataset, (Herranz, L. et al. 2016) proposed restaurant context-based food image recognition which adds additional GPS information with restaurant's menu databases. There are also some regional food datasets, such as (Matsuda, Y. et al. 2012) published UECFOOD100 Japanese food dataset with a bounding box in each image. There are many food datasets and the researchers need to select datasets suitable for the research objectives to train the model.

Some researchers also invested in large-scale food-related study. (Chen, S. et al. 2016) proposed a data-driven fruit counting algorithm that applied a novel deep learning approach trained from large human-generated labels datasets. They used a fully convolutional network to extract each region and applied a second convolutional network to estimate the counts of fruits. (Hassannejad, H. et al. 2016) fine-tuned Google's image recognition architecture Inception using a deep convolutional neural network to classify the food images on three large benchmark datasets and achieved the best results. (Pandey, P. et al. 2017) developed a multilayered convolutional neural network to recognize the contents in the meal images. They trained and fine-tuned CNN with handcrafted features and methods and tested on the largest real-world food recognition database ETH Food-101 and got a comparable result.

Recent efforts on food recognition algorithms based on single-dish photos are not useful for actual needs on multi-dish smart applications. (Ege, T. et al. 2018) proposed CNN-based multi-task learning for detecting multiple-dish and automatically estimate food calories simultaneously. They also annotated the bounding box of multiple dish images and a single dish with calories. This algorithm does not consider the influence of food volume on the calculation of calories. (Hoashi, H. et al. 2010) applied multi-modal feature fusion method with Multiple Kernel Learning for food image recognition system. They developed an automatic food recognition system for fusing image features, such as bag-of-features, Gabor features and gradient histogram. The experiment is conducted on 85 food categories. (Horiguchi, S. et al. 2018) developed an efficient personalized system for the incremental classifier. They implemented a nearest class mean classifier and 1-nearest neighbour classifier on daily food images. They considered solving the limitation of samples for the individual user. The sample contents are different and the amounts of the datasets per person are very small. (Kongsorot, Y. et al. 2014) combined multi-label classification ELM with Canonical Correlation Analysis (CCA-ELM). They first computed the relationships between target labels and input features, then mapped them into new space. Then applied ELM to classify them and map them back to the original space.This CCA-ELM method improves the performances of ELM on multi-label learning classification and recognition.

2.2.2 Food Ingredients Classification

Food ingredients classification is different from food classification. The food classification task usually divided into two types. One is to identify food categories on a picture, for an example, we need to categorize many kinds of food on a picture. The other is to tag the food with different tags, or we would like to identify possible tags for the image. Such as French Fries, which can contain these tags, crabs, fats, appetizer and junk. Food ingredients classification is to identify the ingredients of food and then do classification tasks, such as a burger label contains bread, vegetables, meat and sesame.

Food ingredients identification has broader requirements. For example, if you want to measure the calories of food, you need to identify different food components in order to estimate the approximate calories more accurately. In addition, food recipe recommendations also need to accurately identify the ingredients of the food and measure the nutritional value. (Maruyama, T. et al. 2012) proposed a real-time food ingredients recognition system for recipe recommendation. They developed an app on a smartphone that can allow people to search for cooking recipes by pointing the phone camera to food instantly. However, this research only on 30 kinds of food ingredients which is far from enough. Similarly, (Su, H. et al. 2014) investigated on analysing the correlation between recipe cuisines and ingredients. They applied automatic cuisine labelling technique by associative classification and SVM methods to find the similar cuisines and essential ingredients for the cuisine.

In recent years, there are many researchers develop research in food ingredients classification. (Zhu, Z. et al. 2021) proposed two types of food ingredients classifier by deep learning methods and achieved high accuracy. The authors proposed salient ingredients classifier and segment-based classifier on three novel datasets and trained on Resnet 50 by transfer learning. (Chen J. et al. 2016) proposed deep-based multitask learning architecture and a double output model for simultaneous multi-label food ingredients recognition and food categorization. In addition, they also applied conditional random field (CRF) to deal with context information of ingredients. But their model cannot be generalized to unseen recipes and food types. They published a standard

Chinese dataset for research and the models will be tested on this dataset. (Pan, L. et al. 2017) applied an automatic classification framework called DeepFood for multi-class food ingredients classification. They used transfer learning for deep features extraction and develop a classifier for each of these feature set. They integrated ResNet features, Information Gain feature selection and Sequential Minimal Optimization classifier together and got a high performance of classification. (Chen J. et al. 2017) proposed multi-task learning to recognize food visible and procedural attributes, such as food ingredients, colour, shape, cutting and cooking in a mix-dish food image. They applied an efficient region-wise pooling method on the feature grid, which significantly improves the results. Different from multi-dish recognition, (Wang, Y. et al. 2019) present mixed dish recognition in that they recognize overlapped food in one dish by region level multi-scale methods. Because different food ingredients tend to overlap, they combined the Negative Sampling and targeted pre-trained model to achieve good results on two real data sets.

Convolutional Neural Networks (CNN) has an excellent performance in image feature extraction and is widely used in the field of food image recognition. (Ferra A et al. 2017) regards the problem of food ingredient identification as a multi-label learning problem, and takes CNN with good classification performance as a multi-label predictive learning recipe ingredient list, which can be better promoted to the new data due to its highly variable formula and ingredients. Both (Wang, X. et al. 2015) and (Bolaños et al. 2017) have done in-depth research on recipe recognition. (Wang, X. et al. 2015) used multi-modal data which combined textual information with image feature for recipe recognition. They embedded textual information with each image and got representations from these multimodal data by using SIFT and Bossanova Image Pooling. They applied convolutional neural network deep features for image recognition, while for textual features representation, they used Term Frequency–Inverse Document Frequency (TF-

IDF) in one semantical space to measure the importance of a word. After got the representations they used a pre-trained model Over-Feat feature extractor to get a deep feature of the food. After the whole process, they merged the image deep features with TF-IDF scores by late fusion. Finally, they developed a web search engine for users to identify the relevant recipes by sending food images. (Bolaños et al. 2017) regards food ingredients recognition as a multi-label learning problem, and then uses CNN to learn the list of ingredients in each picture. They use InceptionV3 and ResNet50 as the basic architecture of the model and then use the ILSVRC dataset to pretrain the model. In the last layer of the model, the Softmax function is used for multi-label classification over N possible outputs. This model has achieved very promising results, but it cannot do incremental learning multi-label classification. (Yanai, K. et al. 2015) also used the CNN network to do food ingredients recognition, the difference is that they pre-trained the Deep-CNN model on ImageNet to extract deep features and fine-tuned on benchmark food datasets for simultaneous recognition of food ingredients and categories. As a comparison, (Zhang, X. J. et al. 2016) proposed Deep-CNN multi-task learning algorithm which is able to recognize food dish types, food ingredients and cooking methods simultaneously. They manually built up a dataset from internet and combined three tasks together for training. Some low-layers were shared in the network architecture and got higher accuracy at the end.

Many scholars synthesize a variety of information to identify food ingredients. (He, H. et al. 2015) proposed a Diet Cam for food ingredients detection and classification, they use multi-view multi-kernel SVM to classify the food ingredients based on shape and texture verification model. (Lasod A et. Al 2017) connected artificial neural network (ANN) to the errands of food location and recognition through parameter change which got higher accuracy than a customary strategy. They combined fuzzy logic with ANN to change over the image properties into fuzzy information and distinguish edge pixels in the middle of a Neural Network. (Bolanos, M. et al. 2016) used bounding boxes to identify food types and locate food. They used CNN to train food detection and then use a Global Average Pooling to generate heat maps of food probability. This method achieves high precision and recall score. (Herranz L. et al. 2016) proposed multi-task learning that combined multi-model data using visual images, positional information and external menus knowledge together. The result on six datasets showed that the performance of a task was boosted by integrating these multiple pieces of evidence.

However, the studies mentioned above are all about multi-label classification tasks on fixed datasets. There is almost no incremental multi-label food ingredients classification, and there are only a few papers about incremental multi-class classification, such as (Venkatesan, R. et al. 2016) proposed a progressive learning approach for multi-class classification but not for food ingredients classification. They proposed an algorithm that can adjust hidden layer neurons and output layer neurons when learning new classes in both sequential and simultaneous manner. (Tahir, G. A. et al. 2020) proposed an incremental learning approach for food class classification and gave a scientific and complete analysis on incremental learning. They investigated an open-ended continual learning framework that can dynamically adjusted network architecture to reduce catastrophic forgetting in the process. However, their research is only on multiclass classification, not for multi-label food ingredients classification. The algorithm got state-of-the-art results and my dissertation also inherit some techniques from this paper.

2.3 Extreme Learning Machine and Improvements

(Huang G.B. et al. 2006) proposed the Extreme Learning Machine (ELM), and after that, many scholars invested in the research and put forward more optimization methods. In part 2.2, the study mainly discusses the basic concepts and structure of the Extreme Learning Machine. The definition of ELM and its variants, the advantages and disadvantages of ELM, the optimization methods and its comparison with common machine learning algorithms are introduced. These are the basis for later discussion of improved algorithms. Finally, the study discusses the combinations with multi-label classification.

2.3.1 Principles of Extreme Learning Machines

ELM was proposed as the single-layer feedforward neural networks (SLFNs) because its hidden layer nodes only have one layer, the network structure is shown in Figure 2.1.

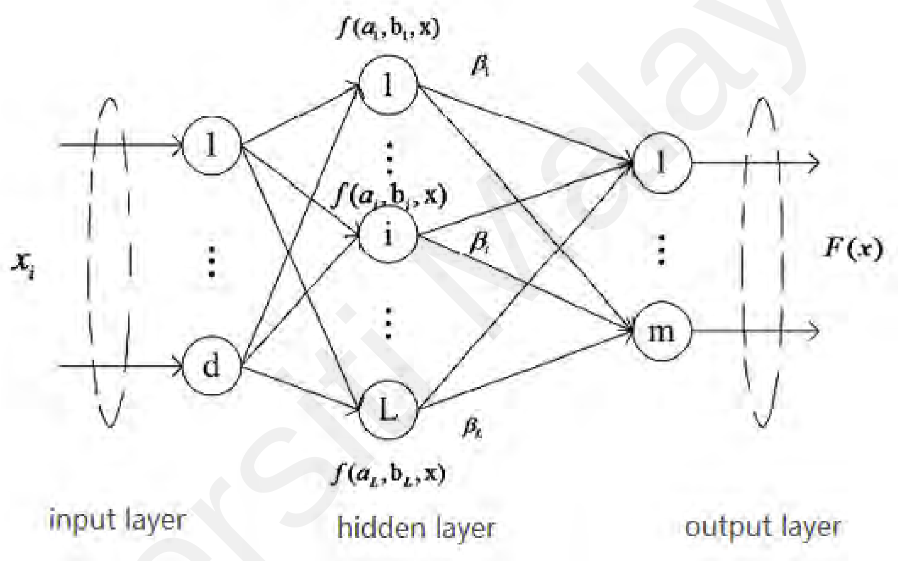

Figure 2.1: Schematic diagram of the ELM

From the figure above, we can see that the entire model structure is divided into three parts: input layer, hidden layer and output layer. The functions of each layer are different. The left is the input layer, it can be used to receive input variables from the environment and pass them to the next layer. The middle layer is the hidden layer, it does not directly receive signals from the outside world, nor send signals directly to the outside world. Its main functions are to realize calculation and identification. The output layer is used to output the results.

As shown in the figure, in the hidden layer, the expression of input weight and threshold value of each node is (a_i, b_i). The $f(a_i, b_i, x)$ is the feature mapping function of each node which is to map the data of the input layer from its original space to the feature space of ELM. The (x_i, t_i) represents the training set, d is the number of input samples and L is the total number of hidden layer nodes. The output weight is β and output function of the hidden layer is:

$$F(x) = \sum_{i=1}^{L} \beta_i f(a_i, b_i, x) \quad (2.1)$$

If O_i represents the learning error of the whole network. Therefore, the entire ELM can be expressed as:

$$O_i = \left\| \sum_{i=1}^{L} \beta_i f(a_{i_1}, b_i, x) - t \right\| \quad (2.2)$$

If H represents the output matrix of the hidden layer, $T = [t_1, \cdots, t_L]^T$, Formula (2.2) will be expressed as:

$$O = \| H\beta - T \| \quad (2.3)$$

When the learning process continues, ELM will gradually reduce the learning error and when reaching its best learning ability, the error is 0, we can get:

$$H\beta = T \quad (2.4)$$

If the hidden layer output matrix H is a non-singular matrix, it can be seen that the number of hidden layer nodes of the network is equal to the number of training samples. In the above formula, we have known H and T, the process of finding the output weight β is also the learning process of ELM. From Equation (2.4), it can be concluded that the solution formula of the output weight β is as shown in (2.5). At this time, the training sample set of the ELM is an error-free fitting.

$$\beta = H^{-1} t \quad (2.5)$$

The H^{-1} is the inverse of H, H represents the irreversible singular matrix. However, in general, the number of training samples is much larger than the number of hidden layer nodes. The learning process of the ELM can be expressed in the following form:

$$\|H\hat{\beta} - t\| = \min_{\beta}\|H\beta - t\| \quad (2.6)$$

The calculation formula of output weight is obtained as follow:

$$\hat{\beta} = H^{+}t \quad (2.7)$$

Where the H^{+} is the Moore-Penrose generalized inverse of H. H^{+} can be obtained by orthogonal projection, iteration, orthogonalization and singular value decomposition. In general, the orthogonal projection method is used. If $H^{T}H$ is a non-singular matrix, we can get to know $H^{+} = (H^{T}H)^{-1}H^{T}$, and if HH^{T} is a non-singular matrix, then $H^{+} = H^{T}(HH^{T})^{-1}$.

(Huang GB. et al. 2011) combined Support Vector Machine and Kernel function, they used Mercer's theorem to construct the Kernel matrix to replace HH^{T} and proposed Kernel Extreme Learning Machine (KELM). In the KELM algorithm, the kernel function is introduced. We do not need to know the feature mapping function F(x) of the hidden layer, nor need to set the number of nodes in the hidden layer. This greatly reduces the computational complexity and improves the generalization ability and stability of the algorithm. The least-square optimization is obtained and the optimization of hidden nodes is avoided. Compared with SVM and basic ELM algorithm, KELM is more stable and has stronger generalization performance. Therefore, KELM is widely used in classification, regression problems.

2.3.2 Extreme Learning Machine and its variants

It can be seen from the above that the network parameters of this algorithm do not need a tedious iterative learning process, and we only need to adjust the number of neurons and regularization parameters, then use the regularization least square method to solve the output weights. ELM needs less computational time with less human intervention obviously but has better generalization performance compared with a traditional neural network. In the general process, for the hidden layer functions, the study can choose any kind of piecewise continuous computational functions, such as Sigmoid, wavelet, fuzzy inference and so on. Just need to predefine the hidden structural parameter nodes and randomly assigned the weights and biases of the inputs. Then calculate the output weights by approximating the minimum norm least-square solution.

Even though ELM has a fast training process, researchers still have invested a lot of researches to improve the speed of computation. (Deng, W. Y. et al. 2016) proposed a fast Reduced Kernel Extreme Learning Machine (RKELM) which reduced the significant cost in the training process, compared with SVM, it only needs a small fraction of the computational task. (Mahmood, S. F. et al. 2017) proposed a fast adaptive shrinkage extreme learning machine (FASTA-ELM) to avoid high time-consuming. They used an extension of forward-backwards splitting to compute the smallest norm of the output weights in ELM. The results on five benchmarked face gender recognition datasets showed that this method had more efficient performance and outperforms. Researchers adopt Cholesky factorization to solve the problem of inversion of the hidden layer output matrix. (Zhou X. R. et al. 2016) applied two comparable schemes of fast online sequential ELM, they are fast online sequential regularized ELM using Cholesky factorization with a forgetting mechanism in the incremental system and Cholesky factorization based online kernelized ELM with forgetting mechanism. They used recursive method for calculating Cholesky factor to reduce the computational burden. Inspired by this method, (Luo, F. et al. 2017) proposed the Cholesky matrix decomposition inverse method for large scale label datasets. They used matrix block method to divide the large matrix into small matrices to calculate the output weights of KELM. The parallel computation of matrix block well solved the large matrix inverse problem. Although research proves that ELM is a high-speed and efficient machine learning method, many kinds of research still optimize it in structure. (Liang N.Y. et al. 2006) proposed Online Sequential Extreme Learning Machine (OS-ELM), which combines ELM and Recursive Least Square that can train new data sample chunk by chunk and meanwhile discard the trained data. Inspired by this method, some researchers developed several variants of this method. (Lan, Y. et al. 2009） put forward an ensemble online sequential ELM (EOS-ELM) algorithm, which is more stable than the original OS-ELM. They integrated several basic OS-ELM frameworks, then used the average output value of these OS-ELM networks as the final measurement of performance. Similarly, (C. Gautam et al 2017) proposed Online Sequential One-Class ELM (OS-OCELM), to overcome the deficiencies of the OCELM that they separate classifiers into reconstruction based and boundary-based with online and offline learning and assigned the threshold function. (Afzal, A. et al. 2021) explored a new multilayer arc-cosine kernel extreme learning machine which has a recursive nature and has the potential to express multilayer computation in learning models.

(Huang G.B. et al. 2006) proposed an incremental extreme learning machine algorithm (I-ELM) which uses a method to determine the number of hidden layer nodes in ELM. In more detail, at first the number of hidden nodes is zero, then they set an error target and continue to add hidden nodes to the network for training. When the increased number reaches the target, the network structure is determined. In addition, (Huang, G. B. et al. 2008) proved that ELM with a complete complex activation function can achieve a lower symbol error rate and provide a simpler structure for complex applications. They applied I-ELM from real domain to the complex domain and randomly added the hidden nodes.

They proved that if the layer activation function is complex continuous discriminatory or complex bounded nonlinear piecewise continuous, the IELM can still maintain similar effect in the complex domain. In contrast to the I-ELM, (Rong, H.J. et al. 2008) proposed Pruning ELM which starts from a large ELM network and set an error target. Then continuously cuts off redundant useless nodes in the network through some methods. (Zhao, Z. et al 2014) proposed a class incremental ELM algorithm for human activity recognition in the cognitive system. They trained a classifier first and then adapt the model to new labelled data. This algorithm cannot be suitable for the unknown data and had less ability for exploring in the real scenario.

2.4 Multiple Classification Methods

In this part, the study mainly discusses some basic concepts and common algorithms of multi-classification. Explains the difference between problem transformation method and the algorithm adaptation method. It also distinguishes the concept of multi-label classification from multi-class classification and multi-task classification. Finally, the study also discussed the algorithms and research of multi-label learning based on Extreme Learning Machined.

2.4.1 Multiple Classification

In real-world applications, one object is usually associated with a set of target labels simultaneously. For an instance, an image of a hamburger containing the labels of bread, eggs, vegetables and meat at the same time. The multi-label learning method belongs to supervised learning which involves predicting multiple output labels for each instance in the training samples (Tsoumakas, G et al. 2007). As mentioned above, one sample can belong to some related labels at the same time, therefore the study cannot simply regard the label with the highest output value as a predicted label. ELM is widely used in regression, clustering, binary classification, multiclass classifications, but seldom in multi-label learning. (Sun, X. et al. 2016) proposed a thresholding method based on ELM, they chose those labels with higher output value than the predefined threshold as the predicted labels. They first trained a multiple classifier based on ELM and then set a threshold function that was learnt from samples by ELM regression. Results showed that their method outperformed in several standard datasets.

In the past few years, many scholars have devoted a lot of researches on multi-label learning. In recent years, there have been many review articles on multi-label learning. (De Carvalho et al. 2009) gave a survey on the comparison and analyzation of multi-label techniques in a pedagogical manner with examples explanations. They used documents classification to illustrate the main techniques and proposed a taxonomy of the similarities and differences in multi-label techniques. (Zhang, M. L. et al. 2013) conducted a survey on multi-label classification from three aspects. They summarized the fundamental concepts and evaluation methods first, and then chose eight representative multi-label learning algorithms to explain in detail. Finally, they also summarized research problems and outlined online resources in this research area.

There are many classification methods for multi-label classification algorithms, and there are two mainstream classifications: one is the problem transformation method and the other is the algorithm adaptation method. The Problem transformation methods convert a multi-label classification problem into a regression problem or one or two single-label classification. The commonly used methods are Binary Relevance, Classifier Chains and Label Powerset. The researchers usually use existing techniques to transform multi-label classification into sub-problems and construct sub-classifiers for each of them. There are a lot of representative bibliography of these kinds of techniques, such as Random k-labelsets (Xu, j. et al. 2012), relevance Binary method (Boutell et al. 2004), Pairwise Multi-label Algorithm (Hüllermeier E. et al. 2008) and so on. (Hüllermeier E. et al. 2008) proposed pairwise comparison ranking method for preference learning. They used a weighted voting strategy and spearman rank correlation to map the instances and minimize the loss function. Results showed that their method is superior in computational efficiency and accuracy. (Xu, j. et al. 2012) proposed simplified Rank-SVM method with adding a zero label for multi-label classification. They divided quadratic programming problems into a series of sub-problems with the Frank–Wolfe method, in which each class has an independent equality constraint. Five novel multi-label classification algorithms were compared, the results showed that their methods had a faster robust performance and less computational cost.

Generally, algorithm adaptation methods are usually to adapt, expand and customize existing specific machine learning algorithms to directly deal with multi-label learning tasks instead of transforming the problem into different subsets of the problem. In addition, the representative algorithms include the RFBoost method (Al-Salemi B. et al. 2016), Ranking Support Vector Machine (Rank-SVM) algorithm (Elisseeff A. E. et al. 2002), etc. To give an example, (Zhang, M. L. et al. 2007) proposed a multi-label k-Nearest (ML-KNN) method to process multi-label data. The model makes inference prediction based on the information embedded in the nearest neighbour and uses Maximum a Posteriori (MAP) to make predictions. (Read J. et al. 2014) applied Monte Carlo schemes for classifier chain in modelling dependencies of correlated classes to solve multi-dimensional classification, both for finding a good chain sequence and performing efficient inference. Their algorithms remain tractable for high-dimensional datasets and obtain the best predictive performance across several real datasets. (Zhang, M. L. et al. 2009） proposed Multi-label Naive Bayes techniques on the instances with multi-labels, which achieved comparable results on real datasets. They also applied principal component analysis for feature extraction and a genetic approach on appropriate feature subset selections. (Al-Salemi B. et al. 2016) improved the traditional

AdaBoost.MH boosting algorithm using ranked features for training, which is called RFBoost. Meanwhile, they applied Labeled Latent Dirichlet Allocation (LLDA) technique for feature selection and reducing feature space. This method is more efficient and effective, the time consuming of the training process is reduced and still have a high score in evaluation methods.

When talking about multi-label classification, two concepts that have to be distinguished are multi-class classification and multi-task classification. The multiclass classification means there are more than two kinds of samples that need to be classified, each sample belongs to one class and only one label, for example, in the age prediction, people can be divided into four categories: children, youth, adult and old people. A person can be in any one class but cannot be both child and old. For the multi-task classification, it means a single estimator has to handle several joint classification tasks at the same time. If go back to the previous example, it will need to predict age section and gender simultaneously.

Multi-Label classification has a very wide range of application scenarios. Many researchers have conducted related research on multi-label classification. Common multi-label classification tasks include text classification, web mining, multimedia labelling, music classification, medical diagnosis, recommendation systems, and scene classification. In semantic scene classification in a photograph (Boutell et al., 2004), the conceptual class may greater than one, including sunsets, cloud and beaches at the same time. A training and testing method for multi-label classification is proposed, and an evaluation matrix including precision score, recall score and the accurate score is adopted. Similarly, (Zhang, M. L. et al. 2006) proposed a back-propagation multi-label learning method, which uses a new error function to capture the features of multi-label learning in text classification and functional genomics problems. The labels belonging to an instance should be ranked with higher probability. After this, the author (Zhang, M. L. et al. 2009) proposed a multi-label adaptive neural network with radial basis function (ML-RBF) to predict the label set about gene functions and natural scene. They first conducted clustering analysis on instances of each possible class, then minimized a sum of squares error function for learning weights. Results on three multi-label datasets showed that their methods achieved highly competitive performance to others.

When dealing with multi-label classification, the researchers also need to consider the relationship between labels. Because these labels are usually not independent that there is a certain correlation among them. There are also many scholars devoted to the correlation algorithm for multi-label learning. (Zhenhai, Z. et al. 2013) applied correlation information entropy on the random k-labels to measure the correlation between labels. They applied information gain to choose subset features and measure the correlation between features and labels. Similarly, (Lee, J. et al. 2016) used conditional entropy and acyclic graph to calculate the correlation of label sets. Although this method got good results, they did not consider the impact of the quality of label sets. Researchers find that construct the features with the labels and labels-to-labels relationship can improve the performance of classification. As mentioned above, (Elisseeff A. E. et al. 2002) proposed Rank SVM using the maximum interval criteria approach to sort loss among relevant labels and achieved good results. However, the complexity of the algorithm increases with the increment of data, so that this algorithm is not suitable for a large number of variables.

In actual scenarios, there are plenty of large-scale multi-label applications with huge numbers of labels. (Yu H.F. et al. 2014) proposed an empirical risk minimization (ERM) framework to handle data with missing labels based on millions of labels. With the continuous expansion of data scale, large-scale multi-label learning has become a research hotspot. For example, some scholars use millions of labels or categories on Wikipedia data to train large-scale classifiers and annotate a new article or web page with a subset of the most relevant categories. Some scholars treat billions of movies as different tags and then recommend a ranked list of labels by movie hotspots to a user in the recommendation system. (Er, M. J. et al. 2017) developed four progressive learning algorithms based on ELM for multi-class classification. They chose Identity Progression, Random Progression, Mean Progression and Historical Progression to compare the performance on several benchmark datasets. However, these datasets have only three to ten different classes.

2.4.2 Multi-label Classification with Extreme Learning Machines

Multi-label learning deals with the problem where each instance is associated with multiple labels simultaneously. The task of this learning paradigm is to predict the label set for each unseen instance, through analysing training instances with known label sets. (Huang, G.B. et al. 2011) proved that ELM provides a feature mapping platform and can be well applied in regression problems and multiple classification applications directly. Many scholars have proposed various algorithms for multi-label learning based on Extreme Learning Machine and their work was specific to a particular task. (Zheng, W. et al. 2013) put forward regularization ELM combined with latent semantic analysis on text categorization. The layer weights can be obtained analytically and achieved a bias-variance balance by adding a regularization term into the linear system. (Kasun, L.L.C. et al. 2013) proposed an ELM Auto Encoder (ELM-AE) classification algorithm, in which inputs are equal to outputs and output weights can be determined to be orthogonal. They used auto-encoders for feature engineering to train multiple-layer neural networks. Furtherly, (Cheng, Y et al. 2019) proposed a kernel extreme learning machine autoencoder (KELM-AE) for a multi-label learning algorithm that fused the feature space information between labels and features and labels correlation. Finally, they tested the approach on multiple published multi-label data sets and found that this algorithm has certain validity and rationality.

Previous researches (Wang, Y. et al. 2019) (Venkatesan, R. et al. 2016) are mostly on the fixed dataset of labels, however, new labels may emerge in data steam in many applications. In order to solve the problem in multi-label learning, researchers have developed a multi-label learning framework based on incremental learning KELM to deal with this problem. (Zhu, Y. et al. 2018) proposed Multi-label learning with Emerging New Labels (MuENL) algorithm to detect and classify the instances with emerging new labels. However, this framework is time and resources consuming as it needs to remodel in each update procedure and the classifiers need iterative tuning. (Leng, Q. et al. 2015) proposed One-Class ELM(OCELM) to identify the new label instances in the single class classification task. They utilized batch theory to train and got fast learning speed, superior predictive performance and good generalization.

Class incremental learning classifier tackles instances with the new emergence of classes in the training procedure. (Wang, T. et al., 2009) proposed a dynamic ELM that can dynamically change the output layer sizes. However, this class incremental mechanism did not work well in multi-label classification with emerging new labels problems. (Kongsorot, Y. et al. 2020) proposed a multi-label incremental kernel extreme learning machine algorithm with newly emerging labels. First, they use a detector to identify instances with new labels and then use a new incremental multi-label classifier to predict the label set of each instance. This classifier can add output units incrementally and update themselves in unmarked instances. However, it is still a user setting framework, which is set through trial and error. (Dave, M. et al 2016) successfully applied incremental learning in multi-label classification and tested on benchmark datasets in diverse domains. They also compared with novel multi-label classifiers, such as SVM, Decision Trees and Nearest Neighbours. The results showed that the algorithm based on ELM is much faster than others and superior for class incremental classification.

Most of the existing researches for food classification used fixed class datasets, cannot learn incremental new classes and new labels. (Horiguchi, S. et al. 2018) proposed personalized classifier adapting to the user's domain incrementally, however, they extracted features from the whole previous data and train the model again. It wastes time and computing resources and makes this approach less favourable. The calculation cost and computational resources associated with retraining are high which makes their approach less favourable. They used a classifier based on ELM instead of using a fixed class softmax classifier, however, the existing methods of ELM are unable to learn continuously.

For this reason, (Tahir, G. A. et al. 2020) proposed a continual learning framework by employing a novel Adaptive Reduced Class Incremental Kernel Extreme Learning Machine (ARCIKELM) for multi-class classification. In the paper, it mentioned that Data incremental learning and Class incremental learning both are important components in incremental learning. Incremental data learning plays a role in improving the performance of existing class recognition and adapting to domain changes by using new images in an open continuous learning system. In contrast to this, the class incremental learning gets knowledge from new classes continuously. Their new novel classifier can learn continuously and adapt the user domain incrementally.

2.5 Summary

In this chapter, the recent years research of Food Recognition, ELM and Multiple Classification Methods are discussed respectively. In the first part, the study discussed the related studies of Food Recognition and Food Ingredients Classification. In the Food Recognition part, the study discussed related algorithms for image recognition and feature selection in recent years, such as ResNet-50, DenseNet, Inception-ResNet-v2 and Relief F method. The ELM-based food recognition and classification algorithms are also discussed.

Then analysed the basic principles of ELM and related algorithms, such as OS-ELM, EOS-ELM, I-ELM, Pruning ELM, FASTA-ELM, RKELM, I-KELM. Then the study also discussed the research related to multi-classification methods and ELM-based multi-label classification. Then used several surveys on the multi-label classification to introduce the basic situation and concepts of multi-label learning. Meanwhile, the study introduced two mainstream classifications methods: the difference between problem transformation methods and algorithm adaptation methods are analysed. Then the research introduced the application scenarios of multi-label classification algorithms and the discussion of label relevance issues. After this, the study also introduced the multi-label classification algorithm based on ELM in recent years and further discussed the application of incremental learning classifiers in multi-label classification.

The study of incremental classification is based on the Kernel Extreme Learning Machine, the difference is it dynamically added the hidden neurons when new labels appeared. The previous study is based on fixed datasets and the algorithm is not suitable for the true scenario because the categories are open-ended.

CHAPTER 3: METHODOLOGY

3.1 Overview

In this chapter, the study introduces the datasets and analyses the general framework of the system from the perspective of methodology. After that, the feature extraction algorithm and Feature selection algorithm are introduced respectively. Finally, the implementation process of the Multi-Label Classification algorithm is described in detail.

3.2 System Overview

As shown in Figure 3.1 below, there are three parts to the framework. The first one is the Feature Extraction module, which plays the role of extracting each incoming food image features during the training period. The second is the feature selection module, the study uses the Relief F method to rank the extracted feature and select the best features based on the proposed strategy. The study uses these features as the final representation of the images. The last one is the Multi-label Classification module. The research proposes a novel adaptive reduced class incremental kernel extreme learning machine called ARCIKELM for multi-label classification.

As shown in the classification part, the red circles are the neurons that are dynamically added to the network. The classifier can incrementally add new output nodes and hidden neurons when the image representations belong to the labels and this will not affect the previous knowledge learnt. If the representations are from existing labels set, it will do not need to adjust new neurons, the system only updates the model sequentially as the setting strategy. It only adds the new hidden nodes when needed. This model satisfies the requirements of open-ended dynamic progressive learning in multi-label classification.

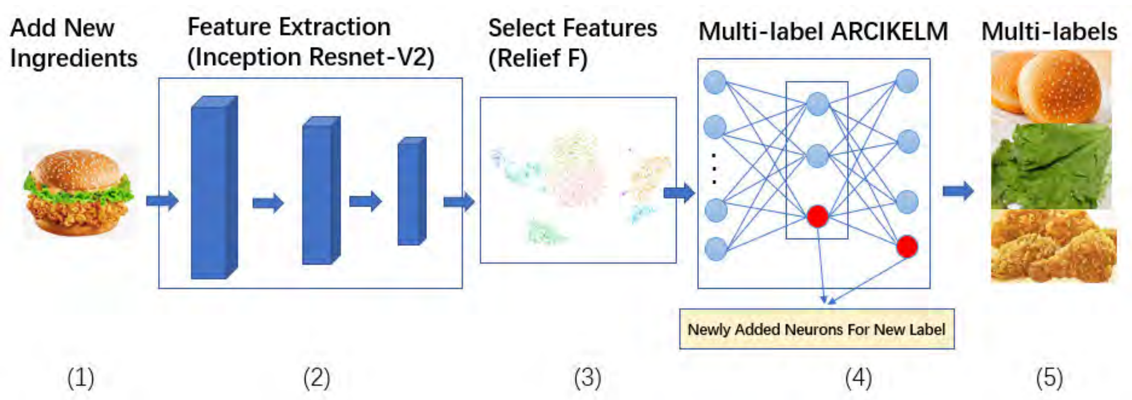

Figure 3.1: Flow chart of proposed framework: The part (1) is the newly added food ingredients. The part (2) is feature extraction part by Inception Resnet V2 network. The part (3) is feature selection part by using Relief F methods. The part (4) is the Multi-label ARCIKELM network, the red neurons is the new nodes. The part (5) is the output multi-labels.

3.3 Datasets Introduction

3.3.1 FOOD101 and Simplified Food101

FOOD101 (Bossard, L et al. 2014) and simplified FOOD101 are both standard benchmarks, the FOOD101 includes 446 labels and Simplified Food101 has 227 labels. They search and collect the latest real-world food dataset which contains 101,000 food images and it was separated into 101 most popular and consistently food classes. Additionally, they sampled 750 images for training and 250 testing images randomly for each class. The side length of images is rescaled to a maximum of 512 pixels while smaller ones were discarded. They deliberately manually clean up the test images, but they did not clean up the training images. Therefore, the training dataset still contains some noise, mostly coming from strong colours and sometimes incorrect labels.

The images folder contains 101 folders, and in each folder, there are 1000 images of a specific food category. Including very diverse but visually and semantically similar food categories, such as Apple pie, Waffles, Bibimbap, Sashimi, Onion rings, Edamame, Macarons to name a few. This makes the FOOD101 is one of the most widely used and most popular datasets for food recognition.

Figure 3.2: Example Images of the FOOD 101

3.3.2 Vireo Food-172 Dataset

Most existing food recognition datasets are either very small in size or have no ingredient level labels. (Chen. J et al. 2016) constructed and released a new dataset called Vireo Food-172. This dataset contains 110,241 food images divided into 172 categories. In addition, they manually annotated the images with 353 visual ingredients labels.

The release of virefood-172 has provided a great boost to the research in the field of food recognition and innovative research on important issues. They crawled the images from both Baidu and Google image search and compiled them by removing duplication. The 172 categories covers eight major groups of foods, including "Vegetables", "Soup", "Bean products", "Egg", "Meat", "Seafood", "Fish", and "Staple".

Figure3.3 : Images of Food Categories in Vireo-Food 172

3.4 Preprocessing Methods

3.4.1 Image Preprocessing

The study also pre-processes the data before train the images, which includes data cleaning and data enhancement. The study mainly uses the method of applying transfer learning with online data augmentation. There are many ways to process images and transform pictures. Common image conversion methods include zoom, horizontal shift, rotation, affine, shear and perspective transformation, etc. The purpose is to add some fresh and credible instances so that the model can better detect features during the training process.

In the experiment, the study preprocesses food images, mainly by shifting, flipping, zooming range, Channel Shift and fill mode and other methods. Specifically, in shifting, the study converts the pixels of the image to the vertical and horizontal directions. In addition, the study mainly does a horizontal flip on the input image by reverse processing the pixels of each row or each column. The study also zoomes in and zoomes out the pictures, because the images come from the real world, and their sizes may come from different levels. At the same time, the study also channel shifting the image data. The red, green and blue values in the pixels are added to the picture through online data augmentation technology, and at last, the image channel is moved by 30 degrees. Finally, in the fill mode transformations, the study uses new pixels to fill the picture boundaries. All the fresh pixels are set to black with value 0 or filled with white with value 1. Table 3.1 is the settings of data transformation.

Table 3.1: Experimental settings of data transformations.

Transformations Types	Values
Wide Shift	0.2
Height Shift	0.2
Horizonal Flip	Random Flip Inputs Horizontally
Zoom Range	0.8 to 1
Channel Shift Range	30
Full Mode	Reflect

3.4.2 Feature Extraction Using Inception-ResNet-V2 Algorithm

After the pre-processing of food images, the study uses a pre-trained deep network on Image-Net to train the enhanced data. Pretrain and fine-tune technology have also been widely used in various fields of artificial intelligence, especially with outstanding performance in image feature extraction. The study loaded large-scale pre-trained models to the existing data through transfer learning, so it did not need to train models from scratch, just fine-tuning, which greatly saved training time. Therefore, the study used this model to greatly reduce the training time and still maintained good recognition results.

(Tahir, G. A. et al., 2020) compared three novel deep learning frameworks, ResNet-50 (He, K. et al., 2015), DenseNet-201(Huang, G. et al. 2017), and Inception-ResNet-V2 (Szegedy, C. et al., 2017) on five different datasets to determine which architecture has higher accuracy and less training time for food recognition. Inception-ResNet-V2 has better performance than ResNet-50 probably because the former integrates the structural advantages of ResNet and Inception model. The Inception-ResNet-V2 can extract features in a fast speed because the residual connection block can accelerate the training. Inception-ResNet-V2 has the best classification results, therefore this study used it as the best architecture for food feature extraction. In the pretrained process, the images are shuffled and random selected. The number of trainable parameters is 54,276,192, after add downstream work, in the finetuned process the parameters increased into 56,211,905.

The original structure and interior of Inception-ResNet-v2 shown in Figure 3.4 below and the parameters tuned in this study shown in Table 3.2 by using Vireo Food-172 datasets as an example.

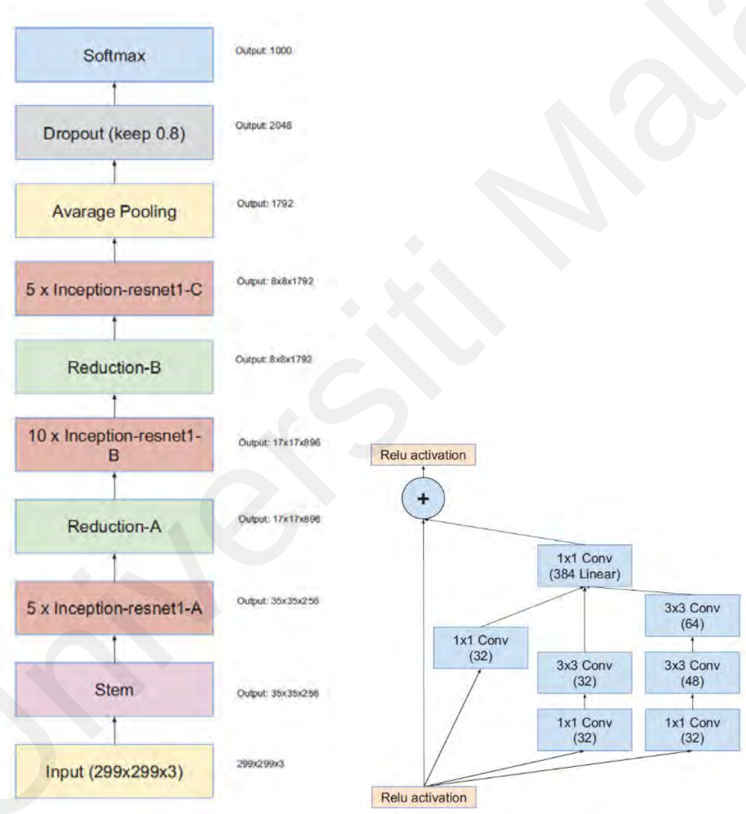

Figure 3.4: Framework of Inception-Resnet-V2 (Szegedy, C. et al. 2017)**:** The left figure is the overall schema of the network and the right figure is interior network structure of Inception-V2, consists of 1x1 and 3x3 convolution. Sizes to the side of each layer summarize the shape of the output for that layer.

(He et al. 2015) proved that Residual connections play an important role in the design of deep network structure. Researchers can design deeper network structures and produce better performance by replacing the filter concatenation with residual connections. For the Inception architecture, it does not need to manually set filter parameters. The network can flexibly and automatically determine these parameters and combine filters. This allows Inception-ResNet-v2 to maintain the original efficient calculation efficiency and simplify the network structure through the advantages of reap residual connections.

Inception-ResNet-V2 is a convolutional neural network that achieves top accuracy on the ILSVRC image classification benchmark. It reduces the amount of calculation by fusing feature maps of different scales and using multiple 3x3 convolutions instead of large convolution kernels. This setup clearly reduces the parameter count by sharing the weights between adjacent tiles. In addition, in Inception-ResNet-V2, the network structure of Resnet and Inception is integrated to further improve the accuracy on ImageNet. Residual connections allow shortcuts in the model, which allows researchers to successfully train deeper neural networks to produce better performance, and also makes it possible to simplify the Inception block.

The study uses the pre-trained ImageNet weights to initialize the model and fine-tuned using food images. Then applies transfer learning on the incremental food datasets to train the deep model which is for extracting features later. The Inception-ResNet-V2 extractor for food recognition can learn image representations well and no need for any handcraft feature extraction. The input shape of each picture is 224x224x3 and the output shape is 353 in the Vireo Food-172 dataset. More detailed composition and changes of parameter dimension of the proposed network architecture shown in Table 3.2.

Table 3.2: The outline of the proposed Inception-Resnet-V2 architecture.

type	input size
conv	224x224x3
conv	111x111x32
conv padded	109x109x32
pool	109x109x64
conv	54x54x64
conv	54x54x80
conv	25x25x192
10x Inception	25x25x128
20 x Inception	12x12x384
10x Inception	5x5x448
pool	5x5x1536
linear	1x1x1024
sigmoid	1x1x353

3.4.3 Feature Selection Using Relief-F Method

Usually, after obtaining the food image feature, the researchers do not use these features directly but choose the optimal feature. Just as in the experiment, the study does not directly use the extracted image feature. Instead, first select a feature subset to remove irrelevant, redundant, and features that do not contribute much to classification. (Tahir, G. A. et al., 2020) proved that not all the features are useful for classification. Relief-F uses convex optimization in an efficient way in weighting features. The weight is calculated depending on the probability of the nearest hits and misses. The scholar (Liu, H. et al. 2007) has confirmed that using a smaller subset of features to describe the dataset is even better than using the original features, and accelerates the learning speed of the algorithm. Therefore, feature selection is an indispensable and very important step in the process of food recognition.

After extracting the features by Inception-ResNet-V2, the study selects the best features for the next processing because of the high dimensionality of the extracted features. The study reduces data dimensionality by removing irrelevant and redundant features, this will reduce the computational complexity during classification. The study chooses to use Relief F (Spolaôr, N. et al., 2013) to rank features and selects the best subset of relevant features with the highest score. Because the Relief F algorithm can be directly applied to deal with multi-label data without any data transformation, so it will not hinder feature dependencies. Finally， the study selected top 500 features for the next step.

There several steps for calculating the estimation of features in a target dataset. The input is a vector of attribute values and the class value, and the output is the qualities vectors. The procedures of Relief F methods are as following:

a) First, randomly select an instance from the extracted features.
b) Second, find k nearest neighbours in the features set, which called k nearest hits.
c) Third, search k nearest features in each different class, which called nearest misses. Conceptually, this encourages the algorithm to estimate the ability of features to separate all pairs of classes regardless of which two classes are closest to one another.
d) Then the study updates the quality estimation for all attributes depending on their values, hits and misses.
e) Finally, in the experiment, the study selected the best 500 features of each instance for the next classification step.

3.5 Adaptive Reduced Class Incremental Kernel Extreme Learning Machine Classification

The last step in the proposed architecture is incremental learning and multilabel classification. In this section, the study will talk about the processing flow of the algorithm and the explanations of multi-label classification. As mentioned in the literature review part, many existing studies are based on fixed food class datasets, which makes the approaches unsuitable for real-time scenarios. The proposed ARCIKELM can learn new labels continuously and can add new hidden neurons when required, which makes the classifier suitable for continual food ingredients recognition.

3.5.1 Classification and Visualization of Label Features

In the final classification stage, the study uses the same deep feature extractor to extract features from the test image dataset. Sort the features by the Relief F method and select the best representations. Although the features extracted from the deep model have good generalization ability, softmax is a fixed class architecture which makes it unable to take advantage of these features. Finally, the study uses ARCIKELM to make the final multi-label classification.

The food category is corresponding to the label of the column with the maximum value of Y and uses a multi-class approach. However, for ingredient detection, the input food representation contains multiple ingredients (Labels). The study uses a trivial threshold of '0' for detecting ingredients labels, corresponding to column Y. The set of columns with values greater than zero gives the belongingness of the corresponding input. For testing multi-label classification, compute the final classification results by using the threshold value. Select the indexes in the Y vector with a threshold greater than zero.

The study also visualizes the increasable labels in the experiment. In the field of machine learning, scholars often use many methods to visualize features in order to observe feature transformations and parameter weight changes more intuitively. Common ones include t-SNE visualization, Matplotlib and Streamlit. The study uses t-SNE tools which is means T-distributed stochastic neighbour embedding to visualize the high dimension features of food label. The study selects four groups, six groups and eight groups of food from the Food101 dataset. The figure 3.5 shows the increasing labels, the leftmost figure represents 4 sets of labels, the middle represents 6 unique group of labels and the rightmost image means 8 unique groups of labels.

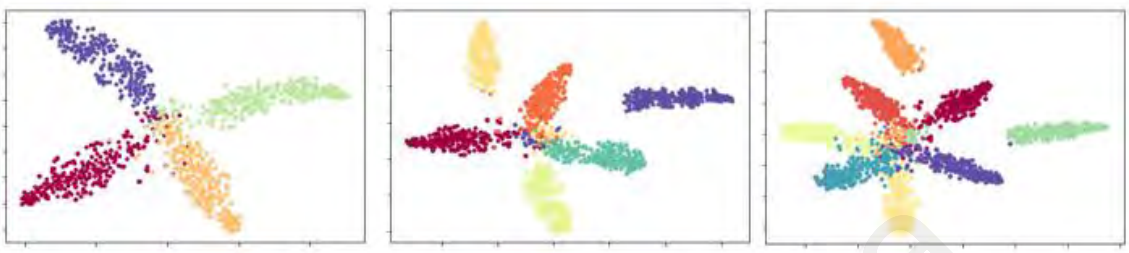

Figure 3.5 Visualization of Increasing Food Labels: Different colours means different labels. This is the process of visualisation of the increasing label when new food ingredients labels were recognized.

3.5.2 The Workflow of the ARCIKELM Algorithm

The workflow of the ARCIKELM classification algorithm includes three phases, the initialization phase, adding new ingredients phase and sequential learning phase. In order to better understand the process, we can also refer to figure 3.4 and figure 3.5. The procedure details are as below:

1. Initialization Phase

First is to select small chunk of training data $\mathbb{N}_b = \left\{\{x_i, t_i\}\right\}_{i=1}^{N_0}$ (3.1), from the all given labels set $\mathbb{N} = \left((x_i, t_i) x_i \in R^m, t_i \in R^l, i = 1,2,3, \dots\dots\right)$ (3.2) to initialize the model.

1. 10% of labels are randomly selected from base labels, marked as X_b
2. Initialize kernel matrix by using

$$k_b = K(X_b, X_b) \quad (3.3)$$

3. Calculate the initial output weight by using

$$\beta^{(b)} = G_b K_b T_b \quad (3.4)$$

4. Set k to 0
5. Initiates the dynamic Neuron Eligibility Matrix N

2. Adding new Ingredients Phase

6. Present the initial new ingredients data
7. Randomly choose 10% of new ingredients Xn
8. To add new output neuron, the system needs to update existing β by use of

$$\beta^{(n)} = \beta^{(b)} \cdot M \quad (3.5)$$

Where M is the information matrix

9. To add hidden neurons for new ingredients Xn by using

$$P_{11} = P_n + P_n K_n^T Z_n S^{-1} Z_n^T K_n P_n \quad (3.6)$$

$$P_{12} = -P_n K_n^T Z_n S^{-1} \quad (3.7)$$

$$P_{21} = -S^{-1} Z_n^T K_n P_n \quad (3.8)$$

$$P_{22} = S^{-1} \quad (3.9)$$

$$U = \beta^{(e)} + P_n K_n^T (I + Z_n S^{-1} Z_n^T (K_n P_n K_n^T - I)) \quad (3.10)$$

$$D = -S^{-1} Z_n^T K_n (\beta^{(e)} + P_n K_n^T (T_n - K_n \beta^{(e)})) + S^{-1} Z_n^T T_n \quad (3.11)$$

Where $P_n = G_n^{-1}$ and G_n is $S = Z_n^T Z_n - Z_n^T K_n P_n K_n^T Z_n$ (3.12), for the detailed derivation process of the formula, researchers can refer to the paper (Tahir, G. A. et al., 2020).

10. To update component G and the kernel matrix Kn

11. To expand the matrix N

3. Sequential Learning Phase

12. To present the (k+1) th chunk of new observations.

$$\mathrm{N}_{k+1} = \left\{(x_i, t_i)\right\}_{i=(\sum_{j=0}^{k} N_j)+1}^{\sum_{j=0}^{k+1} N_j} \qquad (3.13)$$

Where N_{k+1} means the number of observations in k+1 chunk of data

13. To calculate the fuzzy membership of input vector and kernel matrix by using

$$\mu_{ih} = \exp\left(-\frac{|x_i - C_h|^2}{2\sigma^2}\right) \; i = 1, \dots, N; h = 1, \dots L \qquad (3.14)$$

14. The calculate the maximum membership value by formula:

$$h_0 = arg \max_{1 \le h \le H} \left|\mu^h\big(X(k)\big)\right| \qquad (3.15)$$

15. If the h_0 is smaller than a threshold, the system will adjust new hidden neuron by using the same methods as in step 9.

16. Otherwise calculate kernel matrix for the (k+1) th chunk of data N_{k+1} and set T_{k+1} by use of

$$k_{k+1} = k(X_{k+1}, X_L) \qquad (3.16)$$

$$X_{k+1} = \{X_i\}_{\left(\Sigma_{j=0}^{k} \; N_j \;\; +1\right)}^{\Sigma_{j=0}^{k+1} N_j} \qquad (3.17)$$

$$T_{k+1} = \begin{bmatrix} t^T \sum_{j=0}^{k} N_{j+1} \\ \vdots \\ t^T \sum_{j=0}^{k+1} N_{j+1} \end{bmatrix} \qquad (3.18)$$

17. Update component G and compute the output weights βk+1 by using

$$G_{k+1} = G_k - G_k K_{k+1}^T (I + k_{k+1} G_k K_{k+1}^T)^{-1} K_{k+1} G_k \quad (3.19)$$

$$\beta^{k+1} = \beta^{(k)} + G_{k+1} K_{k+1}^T \left(T_{k+1} - K_{k+1} \beta^{(k)}\right) \quad (3.20)$$

18. Set k = k+1
19. If new ingredients labels appear, go to step 6 else go to step 12.

3.6 Summary

In this chapter, the study mainly discusses the overall process and system framework of the system. Then the study introduces three standard benchmark datasets, Food101, Simplified Food101 and Vireo Foods-172, along with the attributes of the dataset, such as data size, characteristics, food categories, number of labels, source of images, etc. Besides, the study implements some preprocessing methods on food images, such as shifting, horizontal flip, zooming, channel shifting. Then two image processing algorithms, Inception-Resnet-V2 for image features extraction and Relief F for selecting features are discussed. Finally, the study describes the details of the algorithm implementation process of ARCIKELM for multi-label classification. The study also explaines the detailed process of multi-label incremental classification and also visualizes the increasing food labels using t-SNE methods.

CHAPTER 4: EXPERIMENTS AND RESULTS

4.1 Overview

In this paragraph, the study mainly introduces experimental design and the implementation environments. The experiment investigates the performance of a state-of-the-art Multi-label ARCIKELM classifier on the three datasets and comprehensively evaluates with F1 score, recall score, Precision Score and Hamming Loss. The study also analyses the experimental results presented in a table, evaluates the performance of the algorithms on three standard data sets and finally visualizes the classification results in the image data. Finally, the study compares the results and gives a brief discussion and explanations on the experiment and results.

4.2 Experimental Setup

4.2.1 Experiment Design

The pre-training and fine-tuning model have gradually become popular in deep learning. The study uses the pre-trained model parameters on ImageNet as initialization parameters and applies them to the different datasets respectively. This study also uses Resnet-50 and InceptionResnetV2 to compare with the ARCIKELM-ML network in downstream work multi-label classification and the evaluation methods are described in detail in the next section. To evaluate the predictive performance of the three novel deep learning networks, the study carries out the experiments on the three benchmark datasets on Google Colab (12 GB RAM, 12 GB GPU). Colab is a free Jupyter Notebook environment and the datasets are stored in the google drive. The version of Python is 3.6 and just need to write and execute code through the browser instead of using local software.

4.2.2 Performance Measurements

In machine learning evaluation standards, there are many methods for results measurement, such as Accuracy, Precision Score, Recall Score, F1 Score, Hamming Loss, etc. Before introducing them, we may need to understand the following concepts in advance.

When doing prediction experiments, each sample can only fall into one of these four scenarios. There is no other possibility. P denotes positive samples and N denotes negative samples. In the first case, when the sample itself is positive and the prediction result is also positive. That means the prediction result is correct, researchers can mark it as True Positive (TP). Second, if the study predicts a sample to be positive, but in fact, the sample is not positive, researchers can mark it False Positive (FP). Third, the sample is positive but the study predicts it as negative, this called False Negative (FN). Last but not least, the actual sample value and the predicted result both are negative, researchers can mark it as True Negative (TN).

In this section, accuracy is also introduced here but the study does not use it in the evaluation, instead, the study uses Hamming Loss. The details are discussed in the Section 4.4 Discussion. Therefore, let us go through Precision Score, Recall Score, F1 Score and Hamming Loss one by one.

The first thing to talk about is the Precision Score, which represents the ability of the model to correctly classify positive samples. In other words, precision also reflects the ability of a model to discriminate N/P samples. The higher the precision score, the stronger the model's ability to distinguish negative samples. The specific formula is:

$$Precision = \frac{TP}{TP + Fp}$$

The second one is Recall Score, in the experiment it represents the ability of a classification model to correctly classify P/N samples among all actual food labels. Recall score reflects the ability of a classification model to identify positive samples, in another word, how many positive samples of the original samples are correctly predicted. There are two situations, one is to predict the original positive labels as positive labels (TP), and the other is to predict the original positive labels as negative labels (FN), so the formula of recall score is as follow:

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

F1 Score is the harmonic average of precision score and recall score. It is also known as the balanced F Score and widely used in the field of information retrieval to measure the performance of retrieval classification and document classification. It also calculates the false positives and false negatives predicted by the model. The value range is between 0 and 1, the higher the F1-score, the more robust the classification model. The F1 score formula is as follows:

$$F_1 = \frac{2 * (Precision * R_{ecall})}{Precision + \mathrm{Re}\,call}$$

Hamming loss is the fraction of the number of wrong labels to the total number of labels. It is used to measure the misclassification of a sample on a single label. In detail, the relevant label does not appear in the predicted label set or the irrelevant label appears in the predicted label set. The optimal value of this loss function is zero and its upper bound is one. The smaller the value, the better the system is. If the loss is 0, it means that all labels of each data have been paired.

$$h = \frac{1}{p} \sum_{i=1}^{p} h(x_i) \Delta Y_i$$

4.3 Experimental Results

a) Record of Results

Table 4.1: Results of each Measurement Method on three datasets.

Dataset	Precision	Recall	F1 Score	Hamming Loss
Food101(Simplified)				
Resnet-50	72.17	68.53	70.3	0.022
InceptionResnetV2	76.1	76.41	76.25	**0.018**
ARCIKELM-ML	**78.36**	**79.12**	**79.04**	0.019
Vireo Food				
Resnet-50	70.381	70.931	70.655	0.005
InceptionResnetV2	70.402	70.948	70.674	0.005
ARCIKELM-ML	**74.345**	**75.291**	**75.364**	**0.0048**
Food101				
Resnet-50	83.423	59.345	69.354	**0.008**
InceptionResnetV2	79.39	76.54	77.94	**0.008**
ARCIKELM-ML	**85.41**	**83.27**	**84.19**	0.011

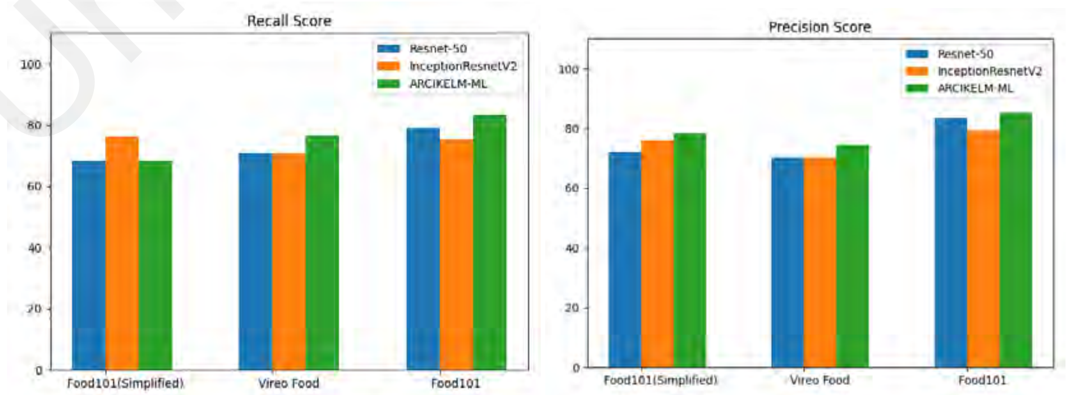

Figure 4.1: Results Comparison Histogram: Comparison and visualization of Precision Score, Recall Score, F1 Score and Hamming Loss on each dataset.

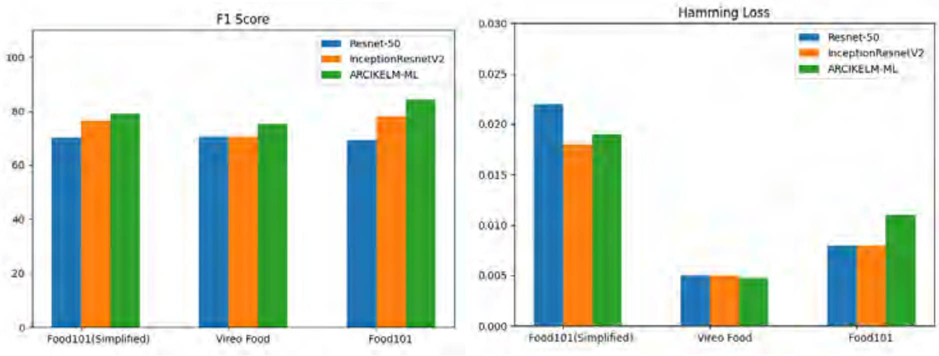

Figure 4.1， Continued

b) Classification Results Visualization

Dish: Braised Sea Cucumber with Scallion
Prediction: 'Minced green onion', 'Salad', 'Meat balls'
Ground Truth: 'Minced green onion', 'Lettuce' 'Meat balls'

(a)

Dish: Streamed twisted rool
Prediction: 'Brunoise diced carrot', 'Chili oil', 'Brunoise diced cucumber', 'Salad dressing'
Ground Truth: 'Brunoise diced carrot', 'Corn Kernels', 'Brunoise diced cucumber', 'Salad dressing'

(b)

Dish: Fried Calamari
Prediction: 'kosher salt', 'all-purpose flour', 'vegetable oil', 'lemon', 'mayonnaise', 'parsley leaves', 'masa harina', 'calaman', 'parsley ', 'meyer lemons'
Ground Truth: 'kosher salt', 'all-purpose flour' 'vegetable oil', 'lemon', 'mayonnaise', 'parsley leaves', 'masa harina', 'calaman', 'parsley ', 'meyer lemons'

(c)

Figure 4.2: Multi-label visualized food ingredients: the example pictures from Vireo Food datasets (a) (b) and Food 101 (c), True Positive labels in green and False Positive labels in red.

4.4 Discussion

ELM and multi-classification are widely used in production, life and research, but there are still many problems that need to be solved and optimized. The research combines the knowledge of the Extreme Learning Machine and incremental learning. This enables the network structure to dynamically increase neurons in the process of machine learning to continuously process newly emerging data. Finally, the study successfully applies the theory to the identification and classification of food ingredients and achieved excellent performance.

To explore answers in the research questions, the experiments are conducted on three benchmark datasets. First, the study pre-processes the image to enhance the data and prepare for the subsequent feature extraction task. Then the study uses Inception-Resnet-V2 for feature extraction. This network structure has an excellent ability to distinguish and identify features. We use transfer learning instead of training from scratch, which saves a lot of training time and computing resources. Through transfer learning combined with incremental classifiers, the model can better and faster distinguish feature extraction features. The study does not use all of the extracted features, instead, ranks features and selects the best representations. At last, the study chooses 500 features of each instance. This removes redundant and useless features, increases the calculation speed and reduces the dimensions of the features. Finally, the study proposes ARCIKELM-ML, which combined Extreme Learning Machine and incremental learning for the multi-label classification task. This network structure can automatically increase hidden neurons and output neurons, and can incrementally learn new food labels. During the entire training process, the study uses four common machine learning evaluation matrices to measure the performance of the model: F1 score, Precision, Recall and Hamming Loss. The final experimental results show that the proposed approach has competitive classification performance besides satisfying the criteria of open-end continual learning.

Feature selection has an important influence on subsequent multi-label classification results. In the result table, on each data set if the feature selection score is high, then the subsequent classification will obtain better results. For example, the F1 score in the third column. In the simplified Food101 dataset, the feature selection score is 76.25 and the multi-label classification score is 79.04. But in the Vireo Food dataset, the scores are

70.674 and 75.364 respectively. When the feature selection score reaches 77.94, the score of multi-label classification rises to 84.19, as shown in the Food101 dataset. This shows that the study must select appropriate and representative features instead of using all features in multi-label classification. It is also necessary to spend more research on feature extraction and feature selection to improve the accuracy of recognition.

For each picture in the test dataset, the study does multi-label classification and label visualization to show the classification more intuitively. In the training set, the study performes image feature extraction, feature selection and multi-label classification tasks. Then records the implementation performance of the three algorithms on the three standard datasets and evaluates them on four metrics which are the Precision score, Recall score, F1 score and Hamming Loss of the three algorithms respectively.

In the first column of Table 4.1 is the precision score, each algorithm has achieved a good performance in each dataset, and the ARCIKELM-ML algorithm has the best performance with a score of 85.41 in Food101 Dataset. The multi-label classification algorithm has the highest scores on the FOOD101 dataset and has a very low Hamming Loss. Just like the introduction of Hamming Loss above, this shows that the algorithm has good classification and recognition capabilities. We also use the Resnet50 network as a comparison. On the three datasets, the evaluation results of Inception-Resnet-V2 and ARCIKELM-ML model perform better than Resnet-50.

The study uses Hamming Loss instead of accuracy because the Hamming distance is suitable for multi-classification problems. It measures the distance between the predicted label and the real label, with a value between 0 and 1. A distance of 0 means that the predicted result is exactly the same as the real result, and a distance of 1 means that the model is completely contrary to the result we want. The experiment demonstrates the feasibility of the incremental multi-classification model in the classification of food ingredients. In the last column, the Hamming loss values are very low. The result of multi-label classification on the Vireo Food dataset is 0.0048, and the scores in the Simplified Food101 and Food101 datasets are 0.019 and 0.011 respectively. This shows that the multi-classification model has a fairly good classification effect on the three standard datasets.

In order to feel the effect of food component recognition and multi-classification better, the study visualizes the experimental results. For each picture in the test set, the study visualizes the food label in each picture and distinguished the correct and incorrect predicted labels by colour. As shown in Figure 4.2, the study shows the ingredients of each food. Green labels indicate correct identification and red colour indicate the wrong classification, such as the red colour in the picture. For each food image, the study has the dish name, prediction results and the ground truth which is the true labels in each dish. For example, in the dish Braised Sea Cucumber with Scallion, the study has a wrong prediction 'Salad' in the results. It may be that the shape of the food is a bit similar to the salad.

4.5 Summary

In this chapter, the study mainly shows the experimental results and analyse and discuss the results. The result data of the three steps of feature extraction, feature selection and multi-label classification are displayed in the table respectively, and visual label processing is performed on each food picture in the test dataset. At the same time, the study also introduces the experimental environment that the study uses the Google Colab online environment, which greatly saves cost and time. Finally, the study describes four classical machine learning evaluation methods and analyse their meanings, thresholds, and formulas.

CHAPTER 5: CONCLUSION AND FUTURE WORKS

5.1 Overview

This is the last chapter of the dissertation, which mainly summarizes the work and results and give a further outlook on the future.

5.2 Conclusion

The research combines Extreme Learning Machine, incremental learning, transfer learning, multi-label classification and food recognition to solve practical food problems. We use the latest deep learning algorithm Inception-Resnet-V2 to extract food features, then use Relief F for feature ranking and selection, and finally use ARCIKELM-ML for multilabel classification. The experiment is based on three standard food datasets, with a total of 1026 labels. The Inception-Resnet-V2 has superior performance in image recognition and feature extraction. Relief F chooses the best features which reduce the feature dimension and subsequent classification time.

In the KELM multi-classification, the continuing growth of data leads to the decline of classification performance. When new samples and labels are added, the model needs to be retrained, which is a huge waste of time and computing resources. Aiming to solve this problem, this dissertation made a study on the multi-label classification based on KELM and applied it to the classification of increasing food ingredients.

The classifier can progressively remodel its structure and can be able to learn newly added labels incrementally in real-time. The experimental results show that the proposed ARCIKELM-ML algorithm is feasible in dynamic multi-label classification tasks and has the characteristics of fast, incremental, transferable, generalizable, high accuracy and continue learning.

5.3 Future Works

There is still a great possibility of improvement in the experimental results. The first is in feature extraction and selection. We think this is the basis for subsequent food recognition and multi-label classification. If there is no good feature extraction and selection algorithm, it will lead to a decrease in the performance of the final recognition and classification. And the experimental results also show that when the feature selection has a high score, the final classification result will be better. Therefore, in the future, the study intends to try to improve the feature recognition and feature selection algorithms. Secondly, the experiment is to do food recognition and classification in the mixed dish data set. In the future, the study can try multi-dish multi-label classification. That is to say, when there are multiple dishes on a picture, the study recognizes different dishes while doing component recognition and multi-label classification for different dishes. Finally, the study can also try to apply it to multi-task food-related research. For example, when the food ingredients are recognized, relevant recipes are recommended to the user at the same time.

In the case of growing food data and unclear data attributes, the algorithm has a good processing effect on the final classification tasks. This increases the confidence in exploring unknown data sets in the future, and the study can migrate the algorithm to exploring new unknown food recognition and labelling tasks. At the same time, some comparative experiments can be added, such as the conventional and latest models that do not use incremental learning.

5.4 Summary

In this chapter, the study gives a conclusion of the whole work in the dissertation, combines with the research objectives and problems, and summarizes the process of experimental treatment and what effect is achieved. Finally, the study discusses the improvement points of the experiments and how to do better in future work.

REFERENCES

Afzal, A. L., Nair, N. K., & Asharaf, S. (2021). Deep kernel learning in extreme learning machines. *Pattern Analysis and Applications*, 24(1), 11-19.

Al-Salemi B., Noah S. A. M, and Aziz M. J. A. 2016. "RFBoost: An Improved Multi-Label Boosting Algorithm and Its Application to Text Categorisation," *Knowledge-Based Systems*, 103:104-117.

Bettadapura, V., Thomaz, E., Parnami, A., Abowd, G. D., & Essa, I. (2015, January). Leveraging context to support automated food recognition in restaurants. *In 2015 IEEE Winter Conference on Applications of Computer Vision* (pp. 580-587). IEEE.

Bossard, L., Guillaumin, M., & Van Gool, L. (2014, September). Food-101–mining discriminative components with random forests. *In European conference on computer vision* (pp. 446-461). Springer, Cham.

Bolaños, M., Ferrà, A., & Radeva, P. (2017, September). Food ingredients recognition through multi-label learning. *In International Conference on Image Analysis and Processing* (pp. 394-402). Springer, Cham.

Bolanos, M., & Radeva, P. (2016, December). Simultaneous food localization and recognition. *In 2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 3140-3145). IEEE.

Boutell M. R., Luo J., Shen X., and Brown C.M. 2004. "Learning multi-label scene classification," *Pattern Recognition*, 37(9):1757-1771.

Bruno, V., & Silva Resende, C. J. (2017). A survey on automated food monitoring and dietary management systems. *Journal of health & medical informatics*, 8(3).

Battiato, S., Caponnetto, P., Giudice, O., Hussain, M., Leotta, R., Ortis, A., & Polosa, R. (2021). Food Recognition for Dietary Monitoring during Smoke Quitting. *In IMPROVE* (pp. 160-165).

Chen, J., & Ngo, C. W. (2016, October). Deep-based ingredient recognition for cooking recipe retrieval. *In Proceedings of the 24th ACM international conference on Multimedia* (pp. 32-41).

Chen, J. J., Ngo, C. W., & Chua, T. S. (2017, October). Cross-modal recipe retrieval with rich food attributes. *In Proceedings of the 25th ACM international conference on Multimedia* (pp. 1771-1779).

Chen, M. Y., Yang, Y. H., ... & Ouhyoung, M. (2012). Automatic Chinese food identification and quantity estimation. *In SIGGRAPH Asia 2012 Technical Briefs* (pp. 1-4).

Chen, M. Y., Yang, Y. H., Ho, C. J., Wang, S. H., Liu, S. M., Chang, E., ... & Ouhyoung, M. (2012). Automatic chinese food identification and quantity estimation. *In SIGGRAPH Asia 2012 Technical Briefs* (pp. 1-4).

C. Gautam, A. Tiwari, and Q. Leng, ``On the construction of extreme learning machine for Online and offline one-class classification - An expanded toolbox,'' *Neurocomputing*, vol. 261, pp. 126_143, Oct. 2017.

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., ... & Kumar, V. (2017). Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robotics and Automation Letters*, 2(2), 781-788.

Cheng, Y., Zhao, D., Wang, Y., & Pei, G. (2019). Multi-label learning with kernel extreme learning machine autoencoder. *Knowledge-Based Systems*, 178, 1-10.

De Carvalho, A. C., & Freitas, A. A. (2009). A tutorial on multi-label classification techniques. *In Foundations of computational intelligence volume 5* (pp. 177-195). Springer, Berlin, Heidelberg.

Dave, M., Tapiawala, S., Er, M. J., & Venkatesan, R. (2016, October). A novel progressive multi-label classifier for class-incremental data. *In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 003589-003593). IEEE.

Dembczyński, K., Waegeman, W., Cheng, W., & Hüllermeier, E. (2012). On label dependence and loss minimization in multi-label classification. *Machine Learning*, 88(1-2), 5-45.

Deng, W. Y., Ong, Y. S., & Zheng, Q. H. (2016). A fast reduced kernel extreme learning machine. *Neural Networks,* 76, 29-38.

Elisseeff, A., & Weston, J. (2001, December). A kernel method for multi-labelled classification. *In NIPS* (Vol. 14, pp. 681-687).

Er, M. J., Venkatesan, R., Wang, N., & Chien, C. J. (2017, November). Progressive learning strategies for multi-class classification. *In 2017 International Automatic Control Conference (CACS)* (pp. 1-6). IEEE.

Ege, T., & Yanai, K. (2018, July). Multi-task learning of dish detection and calorie estimation. *In Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management* (pp. 53-58).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Harper, C., & Siller, M. (2015). OpenAG: a globally distributed network of food computing. *IEEE Pervasive Computing*, 14(4), 24-27.

Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016, October). Food image recognition using very deep convolutional networks. *In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 41-49).

Herranz, L., Jiang, S., & Xu, R. (2016). Modeling restaurant context for food recognition. *IEEE Transactions on Multimedia*, 19(2), 430-440.

He, H., Kong, F., & Tan, J. (2015). DietCam: multiview food recognition using a multikernel SVM. *IEEE journal of biomedical and health informatics,* 20(3), 848-855.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Horiguchi, S., Amano, S., Ogawa, M., & Aizawa, K. (2018). Personalized classifier for food image recognition. *IEEE Transactions on Multimedia*, 20(10), 2836-2848.

Hoashi, H., Joutou, T., & Yanai, K. (2010, December). Image recognition of 85 food categories by feature fusion. *In 2010 IEEE International Symposium on Multimedia* (pp. 296-301). IEEE.

Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks*, 17(4), 879-892.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Huang, G. B., Li, M. B., Chen, L., & Siew, C. K. (2008). Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing*, 71(4-6), 576-583.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2), 513-529.

Hüllermeier, E., Fürnkranz, J., Cheng, W., & Brinker, K. (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172(16-17), 1897-1916.

Kasun, L.L.C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). *Representational learning with ELMs for big data*.

Kawano, Y., & Yanai, K. (2013). Real-time mobile food recognition system. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1-7).

Kawano, Y., & Yanai, K. (2014, September). Food image recognition with deep convolutional features. *In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 589-593).

Knez, S., & Šajn, L. (2015, September). Food object recognition using a mobile device: State of the art. *In International Conference on Image Analysis and Processing* (pp. 366-374). Springer, Cham.

Kongsorot, Y., & Horata, P. (2014, January). Multi-label classification with extreme learning machine. *In 2014 6th International Conference on Knowledge and Smart Technology (KST)* (pp. 81-86). IEEE.

Kongsorot, Y., Horata, P., & Musikawan, P. (2020). An incremental kernel extreme learning machine for multi-label learning with emerging new labels. *IEEE Access*, 8, 46055-46070.

Lan, Y., Soh, Y. C., & Huang, G. B. (2009). Ensemble of online sequential extreme learning machine. *Neurocomputing*, 72(13-15), 3391-3395.

Lasod, A., & Soni, D. (2017, August). Efficiency enhancement of food recognition using artificial neural network. *In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 2758-2762). IEEE.

Lee, J., Kim, H., Kim, N. R., & Lee, J. H. (2016). An approach for multi-label classification by directed acyclic graph with label correlation maximization. *Information Sciences*, 351, 101-114.

Leng, Q., Qi, H., Miao, J., Zhu, W., & Su, G. (2015). One-class classification with extreme learning machine. *Mathematical problems in engineering*, 2015.

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks*, 17(6), 1411-1423.

Liu, H., & Motoda, H. (Eds.). (2007). *Computational methods of feature selection*. CRC Press.

Luo, F., Guo, W., Huang, F., & Chen, G. (2017). Multi-label Learning Based on Kernel Extreme Learning Machine. *DEStech Transactions on Computer Science and Engineering*, (csae).

Manjunathan, A., Lakshmi, A., Ananthi, S., Ramachandran, A., & Bhuvaneshwari, C. (2021). Image Processing Based Classification of Energy Sources in Eatables Using Artificial Intelligence. *Annals of the Romanian Society for Cell Biology*, 7401-7407.

Mahmood, S. F., Marhaban, M. H., Rokhani, F. Z., Samsudin, K., & Arigbabu, O. A. (2017). FASTA-ELM: a fast adaptive shrinkage/thresholding algorithm for extreme learning machine and its application to gender recognition. *Neurocomputing*, 219, 312-322.

Martinel, N., Foresti, G. L., & Micheloni, C. (2018, March). Wide-slice residual networks for food recognition. *In 2018 IEEE Winter Conference on applications of computer vision (WACV)* (pp. 567-576). IEEE.

Maruyama, T., Kawano, Y., & Yanai, K. (2012, November). Real-time mobile recipe recommendation system using food ingredient recognition. *In Proceedings of the*

2nd ACM international workshop on Interactive multimedia on mobile and portable devices (pp. 27-34).

Matsuda, Y., Hoashi, H., & Yanai, K. (2012, July). Recognition of multiple-food images by detecting candidate regions. *In 2012 IEEE International Conference on Multimedia and Expo* (pp. 25-30). IEEE.

Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., ... & Murphy, K. P. (2015). Im2Calories: towards an automated mobile vision food diary. *In Proceedings of the IEEE International Conference on Computer Vision* (pp. 1233-1241).

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019). A survey on food computing. *ACM Computing Surveys (CSUR)*, 52(5), 1-36.

Pan, L., Pouyanfar, S., Chen, H., Qin, J., & Chen, S. C. (2017, October). Deepfood: Automatic multi-class classification of food ingredients using deep learning. *In 2017 IEEE 3rd international conference on collaboration and internet computing* (CIC) (pp. 181-189). IEEE.

Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing foods using ensemble of deep networks. *IEEE Signal Processing Letters*, 24(12), 1758-1762.

Read J., Martino L., and Luengo D. 2014. "Efficient Monte Carlo Methods for Multi-Dimensional Learning with Classifier Chains," *Pattern Recognition*, 47(3):1535-1546.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning*, 53(1), 23-69.

Rong, H. J., Ong, Y. S., Tan, A. H., & Zhu, Z. (2008). A fast pruned-extreme learning machine for classification problem. *Neurocomputing*, 72(1-3), 359-366.

Spolaôr, N., Cherman, E. A., Monard, M. C., & Lee, H. D. (2013, October). ReliefF for multi-label feature selection. *In 2013 Brazilian Conference on Intelligent Systems* (pp. 6-11). IEEE.

Sun, X., Wang, J., Jiang, C., Xu, J., Feng, J., Chen, S. S., & He, F. (2016). ELM-ML: study on multi-label classification using extreme learning machine. *In Proceedings of ELM-2015 Volume 2* (pp. 107-116). Springer, Cham.

Su, H., Lin, T. W., Li, C. T., Shan, M. K., & Chang, J. (2014, September). Automatic recipe cuisine classification by ingredients. *In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication* (pp. 565-570).

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Salim, N. O., Zeebaree, S. R., Sadeeq, M. A., Radie, A. H., Shukur, H. M., & Rashid, Z. N. (2021, July). Study for Food Recognition System Using Deep Learning. *In Journal of Physics: Conference Series* (Vol. 1963, No. 1, p. 012014). IOP Publishing.

Tahir, G. A., & Loo, C. K. (2020). An open-ended continual learning for food recognition using class incremental extreme learning machines. *IEEE Access*, 8, 82328-82346.

Tatsuma, A., & Aono, M. (2016). Food image recognition using covariance of convolutional layer feature maps. *IEICE TRANSACTIONS on Information and Systems*, 99(6), 1711-1715.

Trattner, C., & Elsweiler, D. (2017). Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv*:1711.02760.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3), 1-13.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85, 189-203.

Venkatesan, R., & Er, M. J. (2016). A novel progressive learning technique for multi-class classification. *Neurocomputing*, 207, 310-321.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2285-2294).

Wang, T., Wang, S., & Zhang, H. (2009). Dynamic extreme learning machine: a learning algorithm for neural network with elastic output structure. *In Proceedings. The*

2009 International Symposium on Intelligent Information Systems and Applications (IISA 2009) (p. 271). Academy Publisher.

Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015, June). Recipe recognition with large multimodal food dataset. *In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.

Wang, Y., Chen, J. J., Ngo, C. W., Chua, T. S., Zuo, W., & Ming, Z. (2019, June). Mixed Dish Recognition through Multi-Label Learning. *In Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities* (pp. 1-8).

Xu, J. (2012). An efficient multi-label support vector machine with a zero label. *Expert Systems with Applications*, 39(5), 4796-4804.

Xu, R., Herranz, L., Jiang, S., Wang, S., Song, X., & Jain, R. (2015). Geolocalized modeling for dish recognition. *IEEE transactions on multimedia*, 17(8), 1187-1199.

Yanai, K., & Kawano, Y. (2015, June). Food image recognition using deep convolutional network with pre-training and fine-tuning. *In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.

Yu, H. F., Jain, P., Kar, P., & Dhillon, I. (2014, January). Large-scale multi-label learning with missing labels. *In International conference on machine learning* (pp. 593-601). PMLR.

Zhang, M. L., & Zhou, Z. H. (2006). Multilabel neural networks with applications to functional genomics and text categorization. *IEEE transactions on Knowledge and Data Engineering*, 18(10), 1338-1351.

Zhang, M. L., Zhou Z. H. 2007. "Ml-kNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, 40:2038-2048.

Zhang, M. L. (2009). Ml-rbf: Rbf neural networks for multi-label learning. *Neural Processing Letters*, 29(2), 61-74.

Zhang, M. L., Peña, J. M., & Robles, V. (2009). Feature selection for multi-label naive Bayes classification. *Information Sciences*, 179(19), 3218-3229.

Zhang, M. L., & Zhou, Z. H. (2013). A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8), 1819-1837.

Zhang, X. J., Lu, Y. F., & Zhang, S. H. (2016). Multi-task learning for food identification and analysis with deep convolutional neural networks. *Journal of Computer Science and Technology*, 31(3), 489-500.

Zheng, W., Qian, Y., & Lu, H. (2013). Text categorization based on regularization extreme learning machine. *Neural Computing and Applications*, 22(3), 447-456.

Zhenhai, Z., Shining, L., Zhigang, L., & Hao, C. (2013). Multi-label feature selection algorithm based on information entropy. *Journal of computer research and development*, 50(6), 1177.

Zhou X. R., Wang C. S. 2016. "Cholesky factorization based online regularized and kernelized extreme learning machines with forgetting mechanism" *Neurocomputing*, 174(1):1147-1155.

Zhu, Y., Ting, K. M., & Zhou, Z. H. (2018). Multi-label learning with emerging new labels. *IEEE Transactions on Knowledge and Data Engineering*, 30(10), 1901-1914.

Zhao, Z., Chen, Z., Chen, Y., Wang, S., & Wang, H. (2014). A class incremental extreme learning machine for activity recognition. *Cognitive Computation*, 6(3), 423-431.

Zhu, Z., & Dai, Y. (2021). Food Ingredients Identification from Dish Images by Deep Learning. *Journal of Computer and Communications*, 9(4), 85-101.