A Modified All-and-One Classification Algorithm Combined with the Bag-of-Features Model to Address the Food Recognition Task

Kalliopi Dalakleidi, Myriam Sarantea and Konstantina Nikita
School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou Str., Athens, Greece

Keywords: Diabetes, All-And-One, Bag-Of-Features, Food Recognition System.

Abstract: Dietary intake monitoring can play an important role in reducing the risk of diet related chronic diseases. Automatic systems that support patients to count the nutrient contents, like carbohydrates (CHO), of their meals, can provide valuable tools. In this study, a food recognition system is proposed, which consists of two modules performing feature extraction and classification of food images, respectively. The dataset used consists of 1200 food images split into six categories (bread, meat, potatoes, rice, pasta and vegetables). Speeded Up Robust Features (SURF) along with Color and Local Binary Pattern (LBP) features are extracted from the food images. The Bag-Of-Features (BOF) model is used in order to reduce the features space. A modified version of the All-And-One Support Vector Machine (SVM) is proposed to perform the task of classification and its performance is evaluated against several classifiers that follow the SVM or the K-Nearest Neighbours (KNN) approach. The proposed classification method has achieved the highest levels of accuracy (Acc = 94.2 %) in comparison with all the other classifiers.

1 INTRODUCTION

Diet related chronic diseases, such as obesity and diabetes mellitus, are expanding nowadays. Therefore, an urgent need for dietary intake monitoring arises that can reduce the risk of these diseases. Studies have shown that when patients with diabetes mellitus do significant errors in reporting their dietary intake, there is an increased risk of postprandial hypo- or hyperglycemia. Automatic systems, usually based on a mobile phone, can support patients that suffer from diet related chronic diseases with carbohydrates (CHO) counting. The user first takes a photograph of the upcoming meal with the camera of his mobile phone. Then, the image is processed so that the different types of food are divided from each other and segmented in different areas of the image. A series of features are extracted from each segmented area and are fed to a classifier, which decides what kind of food is represented by each segmented area. Then, the volume of each segmented area is calculated and the total CHO of the depicted meal are estimated.

Feature extraction can play a key role in dietary intake monitoring systems. Efficient feature descriptors could ensure stability and distinctiveness, where stability means that the extracted features are invariant to different photometric and geometric changes and distinctiveness means that the extracted features can be used to distinguish the specified object from other objects or the background. Features related to color and texture have been shown to ensure stability and distinctiveness. Moreover, a large variety of local feature descriptors has been proposed in the literature, like Gaussian derivatives (Florack et al., 1994), moment invariants (Mindru et al., 2004), complex features (Baumberg, 2000; Schaffalitzky and Zisserman, 2002), steerable filters (Freeman and Adelson, 1991), and phase-based local features (Carneiro and Jepson, 2003). A variant of Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), has the ability to capture spatial intensity patterns, while being robust to small deformations or localization errors and is shown to outperform the previous mentioned categories of features (Mikolajczyk and Schmid, 2003; Bay et al., 2008).

Classification results of food images in dietary intake monitoring systems can be improved when the dimension of the extracted feature vector is reduced. The use of the Bag-Of-Features (BOF) model (Peng et al., 2016), which is inspired by the Bag-Of-Words model for text classification (Cruz-Roa et al., 2011) has been reported to highly improve classification accuracy in food recognition tasks. The BOF model achieves dimensionality reduction by creating from

Dalakleidi K., Sarantea M. and Nikita K.
A Modified All-and-One Classification Algorithm Combined with the Bag-of-Features Model to Address the Food Recognition Task.
DOI: 10.5220/0006141302840290
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017)*, pages 284-290
ISBN: 978-989-758-213-4

the extracted features visual words, and by describing the image content with the distribution of these visual words (Wang et al., 2016).

Another important aspect of the food recognition task is that it is usually a multiclass classification problem, as the used food datasets almost always contain more than two categories of food. There exist many classification approaches in order to address the multiclass recognition task, but the most prominent ones, like the One-Against-All (OAA), the One-Against-One (OAO) and the All-And-One (A&O) (Pedrajas and Boyer, 2006) descend from the binarization strategy, where the division of the initial multiclass problem to several binary class problems takes place (Galar et al., 2011).

Several attempts to implement automatic or semi-automatic systems for dietary intake monitoring have been reported in the literature. A food identification application called DietCam has been recently presented (Kong and Tan, 2012), which consists of three parts: image manager, food classifier and volume estimator. Images taken by the users are fed to the image manager, then SIFT features are extracted, clustered into visual words and fed to a simple Bayesian probabilistic classifier, which achieves high levels of accuracy (92%). The food volume estimator calculates the volume of each food item recognized by the food classifier and then the calorie content of the food is estimated. Another food recognition application has been recently proposed for the classification of fast-food images (Shroff et al., 2008). After segmentation of the fast-food image, color, size, texture, shape and context-based features are computed and fed to a feed-forward artificial neural network achieving a 90% accuracy. Moreover, a food identification system has been presented which consists of the following modules: image segmentation, feature extraction, food classification, and volume estimation (Zhu et al., 2010). Food description is based on a set of color and texture features, while classification is based on a Support Vector Machine (SVM) classifier, which has achieved high classification accuracy (95,8%). An automated Food Intake Evaluation System (AFIES) has been reported (Martin et al., 2009), which consists of reference card detection, food region segmentation, food classification and food amount estimation modules. The color RGB data are used as feature vectors for classification, which is performed using the Mahalanobis distance of pixels from food classes. The amount of calorie intake is estimated based on the assumption that food area is linearly proportional to the food volume. In another study, recognition of seven broad categories of food based on a representation for food items that calculates pairwise statistics between local features has been presented (Yang et al., 2010). These statistics are accumulated in a multi-dimensional histogram, which is then used as input to a SVM classifier. Food images are taken from the Pittsburgh Food Image Dataset (PFID) (Chen et al., 2009). This system has also achieved high levels of recognition accuracy (80%).

The use of the BOF model has been adopted in several food recognition systems recently, since food recognition does not presume any typical spatial arrangement of the food elements. Based on the BOF model, the Food Intake Visual and Voice Recognizer system which aims to measure the nutritional content of a user's meal (Puri et al., 2009) has been proposed. Given a set of three images of a user's plate of food, the system first asks the user to list food items through speech, then attempts to identify each food item on the plate, and finally reconstructs them in 3D to measure their respective volumes. Food images are collected by the developers of the system. Food classification is based on the combined use of color neighborhood and maximum response features in a texton histogram model, which resembles the BOF approach. Adaboost is used for feature selection and SVM for classification, which achieves recognition accuracy about 90%. Moreover, a food recognition system for the classification of Japanese food images has been introduced (Joutou and Yanai, 2009), based on the combined use of BOF of SIFT, Gabor filter responses and color histograms, which are then fed to a multiple kernel learning classifier, which has achieved acceptable levels of accuracy (61,34%). The BOF model has been used in another automatic food recognition system (Anthimopoulos et al., 2014). The system firstly computes dense local features using the SIFT on the HSV (Hue Saturation Value) color space, then builds a visual vocabulary of 10000 visual words by using the hierarchical k-means clustering and, finally, classifies the food images with a linear SVM classifier, which achieves high levels of accuracy (78%).

In the present study, a food recognition system is proposed which consists of two modules performing feature extraction and classification of food images, respectively (Figure 1). Motivated by the ability of SURF to capture spatial intensity patterns and the stability and distinctiveness provided by Color and Local Binary Pattern (LBP) features, the combination of SURF, Color and LBP features is examined in this study. Moreover, a novel modified version of the All-And-One (M-A&O) SVM classifier for multiclass classification problems is proposed and its performance is assessed against classification methods based on SVM or the K-Nearest Neighbour approaches including the OAA SVM, the OAO SVM, the A&O SVM, the Weighted K-Nearest Neighbour (WKNN) classifier, the Dual Weighted K-Nearest Neighbour (DWKNN) classifier, and the K-Nearest Neighbour Equality (KNNE) classifier.

2 METHODS

2.1 Dataset

The Food Image Dataset (FID) used in this study consists of 1200 images, 260-by-190 pixels each, collected from the web. Each image belongs to one of six categories corresponding to bread, meat, potatoes, rice, pasta and vegetables (Figure 2). Each category is represented by 200 images. The food is photographed under different servings, view angles, and lighting conditions. The background of every image is edited so that it is completely black.

2.2 Feature Extraction

In the present study SURF, Color and LBP features have been combined to represent each food image in the proposed food recognition system.

SURF detects points of interest using an integer approximation of the determinant of Hessian blob detector, and, then computes the features based on the Haar wavelet response around each point of interest (Bay et al., 2008). Color features are calculated as the average value of color for every 4-by-4 pixel block of the image. LBP is a texture descriptor that provides a unified description, including both statistical and structural characteristics of a texture patch (Prabhakar and Praveen Kumar, 2012). The LBP feature vector is calculated by dividing the image into cells, and comparing the center pixel's value with the neighbours' pixel values of each cell. Then, a histogram of the numbers occurring over the cells is computed. A useful extension to the LBP is the uniform LBP, which reduces the length of the initial feature vector from 256 to 59 (Ojala et al., 2002).

The approach of BOF is used to decrease the input feature space, and deal with high visual diversity and absence of spatial arrangement encountered in food recognition. The BOF approach is influenced by the Bag-Of-Words representation for text classification (Cruz-Roa et al., 2011) and consists of the following two steps. Firstly, a set of small blocks are extracted from each image in the dataset, which are represented by feature vectors. Secondly, the visual dictionary of the image dataset is constructed and each image is represented by the frequency of the codewords of the visual dictionary. The visual dictionary is built with the use of the k-means clustering algorithm. The cluster centers of the feature points extracted in the first step of the BOF approach are defined as visual words. The visual dictionary is the combination of these visual words (Wang et al., 2016).

2.3 Classification

The classification task is performed using a modified version of the All-and-One SVM and its performance is assessed against several classification methods based on the SVM and K-Nearest Neighbours (KNN) approach, including the OAA SVM classifier, the OAO SVM, the A&O SVM, the WKNN classifier, the DWKNN classifier, and the KNNE classifier. All algorithms have been implemented with MATLAB 2015a, are trained with the 70% of the images of the FID, and tested with the rest 30% of the FID.

2.3.1 SVM-based Classifiers

The OAA SVM Algorithm.
The OAA SVM classifier (Galar et al., 2011) consists of K binary SVM classifiers, where K is the total number of classes. The i-th classifier is trained by labeling all the instances in the i-th class as positive and the rest as negative. Each test instance is classified to the class with the biggest score.

The OAO SVM Algorithm.
The OAO SVM classifier (Galar et al., 2011) consists of K(K-1)/2 binary SVM classifiers, where K is the number of classes. Each binary classifier learns to discriminate between a pair of classes. The outputs of these binary classifiers are combined so that the class with the highest score is assigned to the test instance.

The A&O SVM Algorithm.
The A&O SVM algorithm (Pedrajas and Boyer, 2006) combines the strengths of the OAO and OAA approaches. Taking into account that for a high proportion of miss-classifications of the OAA approach, the second best class is actually the correct class, and that the binary classifiers of OAO are highly accurate on their own, but may lead to incorrect results when combined, the A&O approach combines the results of K OAA classifiers and K(K-1)/2 OAO classifiers. The A&O approach first classifies a test instance using the K OAA classifiers and holds the two classes i,j with the biggest scores. Then, the binary classifier of the OAO approach is used to classify the instance among classes i,j.

The M-A&O SVM Algorithm.
The M-A&O SVM algorithm combines the strengths of the OAO and OAA approaches as the A&O SVM algorithm, but in a different way. The M-A&O SVM approach first classifies a test instance using the K OAA SVM classifiers and holds the scores. Then, the K(K-1)/2 SVM binary classifiers of the OAO approach are used to classify the instance. The test instance will be assigned to the class that will achieve the highest score from all (K + K(K-1)/2) classifiers.

2.3.2 KNN-based Classifiers

The WKNN Algorithm.
The WKNN algorithm is a modified version of the K-Nearest Neighbours (KNN) algorithm. According to the KNN algorithm, the k-nearest neighbours of the query instance are selected according to a distance criterion, such as the Euclidean distance. Then, the query instance is assigned to the class represented by the majority of its k-nearest neighbours in the training set. In the WKNN algorithm, the closer neighbours are weighed more heavily than the farther ones (Marinakis et al., 2009) and the distance-weighted function w_i to the i-th nearest neighbor is defined as,

$$w_i = \frac{k+1-i}{\sum_{m=1}^{k} m}$$

where m is an integer in the interval (1,k) and k is the total number of the neighbours.

The DWKNN Algorithm.
In order to address the effect of the number of neighbours on the classification performance, a DWKNN algorithm has been proposed (Gou et al., 2011). The DWKNN algorithm gives different weights to the k nearest neighbours depending on distances between them and their ranking according to their distance from the query object (Dalakleidi et al., 2013). The distance-weighted function w_i of the i-th nearest neighbor is calculated according to the following equation,

$$w_i = \begin{cases} \dfrac{d_k^{NN} - d_i^{NN}}{d_k^{NN} - d_1^{NN}} \times \dfrac{1}{i}, d_k^{NN} \neq d_1^{NN} \\ 1,\ d_k^{NN} = d_1^{NN} \end{cases}$$

where d_i^{NN} is the distance of the i-th nearest neighbour from the query object, d_1^{NN} is the distance of the nearest neighbour, and d_k^{NN} is the distance of the k-furthest neighbour. Thus, the weight of the nearest neighbor is 1, and the weight of the furthest k-th neighbor is 0, whereas other weights are distributed between 0 and 1.

The KNNE Algorithm.
The KNNE algorithm (Sierra et al., 2011) is a variation of the KNN classifier for multiclass classification. It searches for the K-nearest neighbours in each class and assigns the query instance in the class whose K-nearest neighbours have the minimal mean distance to the test instance.

3 RESULTS

The FID is used for the evaluation of the proposed classification algorithm against the classification algorithms based on the SVM and KNN approach on the food recognition task. In order to improve the classification accuracy of the examined algorithms, several sizes of the vocabularies of the BOF model are tested. Table 1 shows the average accuracy of the OAO SVM classifier on the six food classes for different sizes of the vocabulary of the BOF model for SURF and Color features. The size of the vocabularies has been varied from 100 to 2000 words. As it can be observed from Table 1, the lowest accuracy (Acc = 85.0%) is achieved with the size of 300 for both the SURF and Color BOF vocabularies, whereas the highest accuracy (Acc = 93.9%) is achieved with the size of 1000 for both the SURF and Color BOF vocabularies. It is also important to note that among the three types of features, Color features contribute the most to the accuracy of the OAO SVM classifier.

Table 1: Average accuracy of the OAO SVM classifier on the six food classes of Food Image Dataset for varying size of the vocabulary of the BOF model for SURF and Color features.

Features			**Acc**
SURF	**Color**	**LBP**	
100	100	59	87.5
200	200	59	90.0
300	300	59	85.0
400	400	59	90.6
500	500	59	91.1
600	600	59	91.7
700	700	59	92.5
800	800	59	93.1
900	900	59	90.6
1000	1000	59	**93.9**
1100	1100	59	93.3
1500	1500	59	91.4
2000	2000	59	90.8

Table 2: The average accuracy (%) of the classifiers under comparison on the six food classes of the Food Image Dataset.

Algorithm	**Acc (%)**
WKNN	84.4
DWKNN	92.8
KNNE	93.9
OAA SVM	90.6
OAO SVM	93.9
A&O SVM	90.3
M-A&O SVM	**94.2**

Table 3: Confusion matrix of the M-A&O SVM for each food class (Bread, Meat, Pasta, Potatoes, Rice and Vegetables) of the Food Image Dataset.

Confusion Matrix						
Acc (%)	Br	M	Pa	Pot	R	Veg
Br	**93.3**	0.0	0.0	0.0	6.7	0.0
M	1.7	**95.0**	0.0	3.3	0.0	0.0
Pa	0.0	0.0	**93.3**	6.7	0.0	0.0
Pot	0.0	1.7	5.0	**93.3**	0.0	0.0
R	0.0	0.0	6.7	3.3	**90.0**	0.0
Veg	0.0	0.0	0.0	0.0	0.0	**100.0**

In Table 2, the average accuracy of the classifiers under comparison on the six food classes is presented. The size of the vocabulary of the BOF method for SURF and Color features is 1000, thus a total number of 2059 features is used for the classification. Ten k-nearest neighbours are used for the WKNN, DWKNN and KNNE. As it can be observed in Table 2, the lowest average accuracy (Acc = 84.4%) is achieved by the WKNN classifier, whereas the highest average accuracy (Acc = 94.2%) is achieved by M-A&O SVM. The second best average accuracy is achieved by the OAO SVM and KNNE algorithms. The superiority of M-A&O SVM can be explained by the fact that it combines two very powerful strategies, the OAA SVM and the OAO SVM, for multiclass classification.

In Table 3, the classification accuracy of M-A&O SVM for each food class is shown in the form of the confusion matrix. It can be observed that the lower classification accuracy (Acc = 90.0%) is achieved for the class of rice. This is due to the fact that rice is mingled with several sauces which can be very different in color and texture. It is important to note that rice is misclassified to potatoes and pasta which are closer to it in terms of CHO than with meat or vegetables. The best classification accuracy is achieved for vegetables (Acc = 100.0%), this is due to the distinctive green color of vegetables.

4 CONCLUSIONS

Automatic food recognition systems can be used to estimate the content of a meal in CHO for patients with diet related chronic diseases, such as obesity and diabetes mellitus. In this study, an attempt to address the tasks of feature extraction and food image recognition was made. The use of the SURF, Color and LBP features in combination with the BOF model has proven to be particularly effective in terms of average classification accuracy. Several classification approaches for multiclass classification have been tested. The best classification accuracy (Acc = 94.2%) has been achieved by a modified version of the All-And-One SVM approach and is quite high as compared to reported values of classification accuracy for food images in the literature (60%-96%). The proposed system can be combined with an image segmentation module and a volume estimation module towards the development of an automatic food recognition system. Moreover, several other classifiers, like AdaBoost, Random Forests and Convolutional Neural Networks, can be used in the future for comparison purposes in the classification module.

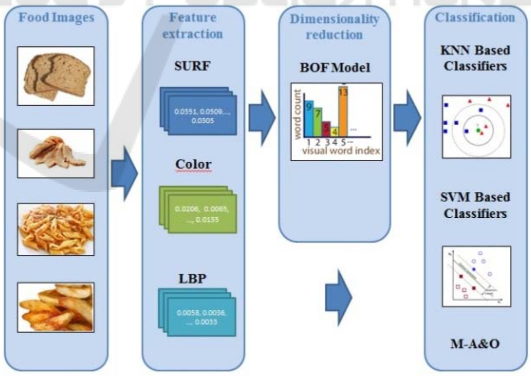

Figure 1: Block diagram of the proposed system.

Figure 2: Example of images from each of the six categories of the Food Image Dataset.

ACKNOWLEDGEMENTS

The work of Kalliopi Dalakleidi was supported by a scholarship for Ph.D. studies from the Hellenic State Scholarships Foundation "IKY fellowships of excellence for post-graduate studies in Greece-Siemens Program".

REFERENCES

Anthimopoulos, M., Gianola, L., Scarnato, L., Diem, P. and Mougiakakou, S., 2014. A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *Journal of Biomedical and Health Informatics*, 18(4), pp. 1261-1271.

Baumberg, A., 2000. Reliable feature matching across widely separated views. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Hilton Head Island, South Carolina, USA: IEEE, pp. 774 – 781.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), pp. 346-359.

Carneiro, G. and Jepson, A., 2003. Multi-scale phase-based local features. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Madison, Wisconsin, USA: IEEE, pp. 736 – 743.

Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R. and Yang, J., 2009. PFID: Pittsburgh fast-food image dataset. In: *IEEE International Conference on Image Processing*. Cairo, Egypt: IEEE, pp. 289-292.

Cruz-Roa, A., Caicedo, J. and Gonzalez, F., 2011. Visual pattern mining in histology image collections using bag of features. *Artificial Intelligence in Medicine*, 52, pp. 91-106.

Dalakleidi, K., Zarkogianni, K., Karamanos, V., Thanopoulou, A. and Nikita, K., 2013. A Hybrid Genetic Algorithm for the Selection of the Critical Features for Risk Prediction of Cardiovascular Complications in Type 2 Diabetes Patients. In: *13th International Conference on BioInformatics and BioEngineering*. Chania, Greece: IEEE.

Florack, L., Haar Romeny, B., Koenderink, J. and Viergever, M., 1994. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4, pp. 171–187.

Freeman, W. and Adelson, E., 1991. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, pp. 891 – 906.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and Herrera, F., 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44, pp. 1761-1776.

Gou, J., Xiong, T. and Kuang, Y., 2011. A novel weighted voting for K-Nearest neighbour rule. *Computers*, 6(5), pp. 833-840.

Joutou, T. and Yanai, K., 2009. A food image recognition system with multiple kernel learning. In: *IEEE International Conference on Image Processing*. Cairo, Egypt: IEEE, pp. 285–288.

Kong, F. and Tan, J., 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, 8(1), pp. 147-163.

Marinakis, Y., Ntounias, G. and Jantzen, J., 2009. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Computers in Biology and Medicine*, 39, pp. 69-78.

Martin, C., Kaya, S. and Gunturk, B., 2009. Quantification of food intake using food image analysis. In: *Engineering in Medicine and Biology Society Annual International Conference of the IEEE*. Minneapolis, Minnesota, USA: IEEE, pp. 6869–6872.

Mikolajczyk, K. and Schmid, C., 2003. A performance evaluation of local descriptors, In: *IEEE Conference on Computer Vision and Pattern Recognition*. Madison, Wisconsin, USA:IEEE, pp. 257 – 263.

Mindru, F., Tuytelaars, T., Van Gool, L. and Moons, T., 2004. Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94, pp. 3–27.

Ojala, T., Pietikäinen, M. and Mäenpää, T., 2002. Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp. 971-987.

Pedrajas, N. and Boyer, D., 2006. Improving multiclass pattern recognition by the combination of two strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6), pp. 1001–1006.

Peng, X., Wang, L., Wang, X. and Qiao, Y., 2016. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Computer Vision and Image Understanding*, pp. 1-17.

Prabhakar, C. and Praveen Kumar, P., 2012. LBP-SURF Descriptor with Color Invariant and Texture Based Features for Underwater Images. In: *8th Indian Conference on Computer Vision, Graphics and Image Processing*. Mumbai, India.

Puri, M., Zhu, Z., Yu, Q., Divakaran, A. and Sawhney, H., 2009. Recognition and Volume Estimation of Food Intake using a Mobile Device. In: *Workshop on Applications of Computer Vision (WACV)*. Snowbird, Utah, USA: IEEE, pp. 1-8.

Schaffalitzky, F. and Zisserman, A., 2002. Multi-view matching for unordered image sets, or “How do I organize my holiday snaps?”. In: *7th European Conference on Computer Vision*. Copenhagen, Denmark: LNCS, pp. 414 – 431.

Shroff, G., Smailagic, A. and Siewiorek, D., 2008. Wearable context-aware food recognition for calorie monitoring. In: *12th IEEE International Symposium on Wearable Computers*. IEEE, pp. 119–120.

Sierra, B., Lazkano, E., Irigoien, I., Jauregi, E. and Mendialdua, I., 2011. K nearest neighbor equality: giving equal chance to all existing classes. *Information Sciences*, 181(23), pp. 5158–5168.

Wang, R., Ding, K., Yang, J. and Xue, L., 2016. A novel method for image classification based on bag of visual words. *Journal of Visual Communication and Image Representation*, 40, pp. 24-33.

Yang, S., Chen, M., Pomerleau, D. and Sukthankar, R., 2010. Food recognition using statistics of pairwise local features. In: *IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, California, USA:IEEE, pp. 2249–2256.

Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C., Ebert, D. and Delp, E., 2010. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE Journal of Selected Topics in Signal Processing*, 4(4), pp. 756–766.