A Counter Medication Classifier Using Machine Learning and Computer Vision Techniques

Eduardo Ceh-Varela[1], Gandhi Hernández-Chan[1], Marisol Villanueva-Escalante[2],
Giner Alor-Hernández[3], José Luis Sánchez-Cervantes[4]

[1] Universidad Tecnológica Metropolitana,
Information Technology and Communication School, Mérida,
Mexico

[2] Instituto Tecnológico de Mérida, Mérida,
Mexico

[3] Instituto Tecnológico de Orizaba, Division of Research and Postgraduate Studies, Orizaba,
Mexico

[4] Professor CONACYT commissioner at Instituto Tecnológico de Orizaba,
Mexico

{eduardo.ceh, gandhi.hernandez}@utmetropolitana.edu.mx,
{mvillanueva}@itmerida.mx, galor@itorizaba.edu.mx, jsanchezc@ito-depi.edu.mx

Abstract. Self-medication and self-prescription are common practices that can be observed in many countries around the world, from the most advanced in terms of medical services in Europe to the less ones as in South America or Africa. Self-medication is defined as the consumption of one or more drugs without the advice of a physician. Many studies in Mexico reveal the type of medications that are consumed as well the social groups that normally use this practice. The consequences for this can range from a mild allergic reaction to death. On the other hand, it is easy to buy drugs without a prescription in pharmacies or supermarkets, but consumers do not always know which one to choose, neither the ingredients nor side effects they can cause. Here we present a classifier model for counter medication based on computer vision and machine learning techniques. We collected 150 images from 11 different counter medications. The classifier was tested with 43 new images, and obtained 90.7% of accuracy, 93% of precision, 91% of recall and 91% of F1-score.

Keywords: Self-medication, computer vision, machine learning.

1 Introduction

Self-medication or self-prescription can be defined as the action of obtain and consume one or more medicines without the advice of a physician nor prescription for a treatment

[1, 2]. As it was stated by the World Health Organization (WHO) and the European Commission of Community Pharmacies, the use of drugs without prescription has been increased due the over-supply and the impact of the media communications concerning self-care to prevent and cure diseases [3, 4]. Studies like the one presented in [5] revealed that in México self–medication problems were observed mostly in women and in adult population with less access to health services. This study also states that the most consumed drugs are analgesics, antibiotics and antacids. Even when self-medication is a common practice, potential risks exist, for example an incorrect self-diagnosis, adverse reactions, incorrect administration or dosage, risk of dependence and abuse, delaying medical advice, and in the worst-case scenario to death [6, 7].

On the other hand, telematics applications have gained success in the field of health promotion, disease prevention and medical treatment. They are attractive because of the potential to provide information adapted to users' needs, social condition and access to technology. The benefits derived from the use of these applications can deeply affect the health status of the population, because they allow citizens to take better decisions regarding their health [8].

In this paper, we present a classifier model that uses machine learning and computer vision techniques to classify counter medications based on an image. We implement this classifier inside a web application that allows users to take a picture of a medication's box in order to receive information related to its content, such as its chemical composition, possible secondary effects, and general comments. The main motivation for this study was to present a tool that can be used to make better decisions about which counter medication to buy.

It is necessary to say that we do not intend to eliminate the problem of self-medication or encourage it, but we pretend to reduce the risks associated. We decided to work with counter medications which can be found in supermarkets, convenience stores or even drugstores.

In the next sections we present the state of the art about self-medication in México and artificial intelligence in the health field, the methodology that we followed for designing and building classifier, the results obtained from the different tests, and finally the conclusions.

2 Self-Medication in Mexico

In reference to Mexico, a study about self-medication in Morelia found that the most frequently purchased drugs were analgesics, anti-inflammatories, vitamins, antibiotics, flu, and benzodiazepines. The study revealed that 51.4% of sales were for self-medication and 13% for recommendation of the pharmacy clerk [9]. A survey applied to 1,537 individuals in Cuernavaca found that the 53.3% consumed drugs by self-medication, of these 64.9% were woman, and 7.6% were between 25 and 44 years old, and the most consumed drugs were analgesics and antibiotics [10].

A total of 245 residents of Tuxtla Gutierrez, in the state of Chiapas, participated in a study where more than half of the participants reported to have taken a medicine without prescription over the last 30 days. Self-medication was significantly more

frequent among older adults who lived alone, and among the illiterate or those with a low level of education [11]. Another survey applied to adults showed that self-medication is related to socioeconomic status and the lack of access to professional healthcare. It also states that the lack of government-sponsored health insurance coverage increases the propensity to self-medicate [12].

A survey applied to 1,859 university students in the state of Puebla showed that 96% of the respondents used drugs without medical advice. It also showed that 97.7% of the students have consumed AINES (acetylsalicylic acid, pseudoephedrine, analgesic, antipyretic) and its combinations, 42.8% mucolytics, 33.6% antidiarrheal, 32.9% antiparasitic, 28.8% antimicrobials, 6.5% topical drugs for acne treatment, and 5.8% oral contraceptives [4].

3   Artificial Intelligence in Medicine

This research turns around Artificial Intelligence (AI), especially in computer vision and machine learning. In this section we present some examples and cases where these areas have been used in medical and pharmaceutical fields.

The earliest work in medical artificial intelligence dates from the early 1970's in the United States, and now is a worldwide field with important contributions from around the globe [13]. Medicine is a field in which AI is critically needed because of the increasing expectations of healthcare [14]. In [15] it is presented a knowledge-based system to increase the overall effectiveness of physicians' time, and thus the quality of healthcare, by improving the information exchange between physicians and patients in clinical settings. The research in [16] focuses on the use of machine learning and cluster analysis techniques to classify a set of organic molecules into their pharmacological activity of sedative and tranquilizer. This was based on the premise that a pharmacological activity of any molecule is dependent on its structure and that structural changes may lead to changes in the activity.

The use of drugs is also a field addressed by AI. Several programs focusing on drug's therapy have been developed. They provide guidance on drug interactions, drug therapy monitoring, and drug formulary selection. There are many aspects of pharmacy where AI can have an impact on [17].

Computer vision is part of the AI field which is centered in theoretical studies of visual information processing. Its two main goals are to develop image understanding systems, and to understand human vision [18]. From its beginning, computer's graphics have been applied to medical fields with research challenges. In modern medicine, medical imaging systems are only one prominent example [19]. In this field, one of the used techniques is known as image matching, which can be considered as a classification scheme than can be used with 2D and 3D images and other signal modalities as electroencephalography and magnetoencephalography [20].

In medicine, the analysis of 3D images can improve significantly diagnosis and therapy [21]. The multimedia workstation of a physician could not exist without tools for manipulating images to perform measurements and to extract and collect
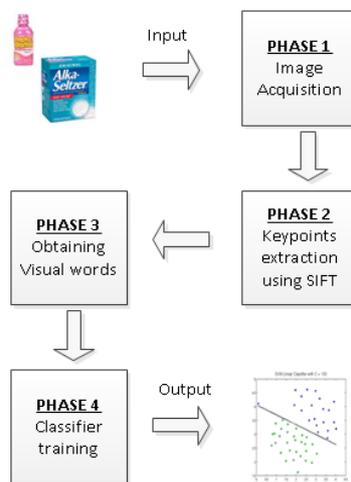
Fig. 1. Four phase methodology.

information. Image analysis and computer vision constitute a wide and rapidly evolving field [22].

Machine learning is used to extract information from the raw data in databases. The process consists on taking the data, and inferring whatever structure in it [23]. Today machine learning provides several indispensable tools for intelligent data analysis, and is currently well suited for analyzing medical data, for medical diagnosis [24]. Another example of the use of machine learning in medical field is the framework proposed by [25]. It is a content-based image retrieval framework for a collection of medical images, and consists of machine learning methods for image prefiltering, similarity matching using statistical distance measures, and a feedback scheme

4   Methodology

We extended the methodology presented in [26] to a four phase methodology as can be seen in figure 1. The first phase was the acquisition of images of counter medications that served to train the classifier. In the second phase the SIFT algorithm was used to extract the images relevant characteristics as keypoints. In the third phase, we used the visual-bag-of-words method to extract a visual word vector for each image; this information was added to the data set to be used in the next step. Finally, the data set was used to train a classifier in order to create a classifier model that identifies new instances of counter medications.

4.1   Image Acquisition

We collected 150 images from 11 different counter medications found on supermarkets and drug stores. The images were collected using the camera of an iPhone 6 and a Sony

Xperia Z2; also we used images downloaded from Internet. Table 1 shows the commercial name for each medication and the number of collected images.

The images were taken with the cell phone cameras in different angles, backgrounds, and lighting. They were JPEG and PNG file format. Also, we group together different boxes of the same medication. An example of these images can be seen in figure 2. We also resized all the images to a same size because of the different images' resolution and convert all to PNG format for better handling.

Table 1. Medication images distribution.

#	Commercial name	Number of images
1	Alka seltzer	11
2	Aspirin	11
3	Bengay	11
4	Bepanthol	11
5	Iliadin	11
6	Pepto bismol	27
7	Sedalmerk	9
8	Sensibit D	11
9	Tempra	19
10	Treda	11
11	Tylenol	18
Total		150

Fig. 2. An example of the images used for training.

4.2   Features Extraction

We use the Scale Invariant Feature Transform (SIFT) method to detect local invariant descriptors inside of each of the medication's images. One of the advantages of using SIFT is that descriptors do not change if the image is scaled, rotated, or even its intensity is changed [31]. It uses the difference-of-Gaussian (DOG) functions to identify potential interest points (keypoints), where these points across both image location and scale are the maxima and minima of (1).

$$D(\pmb{x}, \sigma) = [G_{k\sigma}(x) - G_{\sigma}(x)] * I(x) = [G_{k\sigma} - G_{\sigma}] * I = I_{k\sigma} - I_{\sigma} \qquad (1)$$

This method computes a descriptor for each keypoint based on its position, scale and rotation [32]. For each keypoint a feature vector was created. This vector is formed by a 4x4 keypoint subregions with 8 bins orientation histogram, resulting in a 128 bins histogram (4x4x8).

In order to apply this algorithm to our project we used the library OpenCV 2.4.9. On the left side of figure 3 the keypoints are shown for a medication box, and on the right side is the SIFT histogram for one of those keypoints.
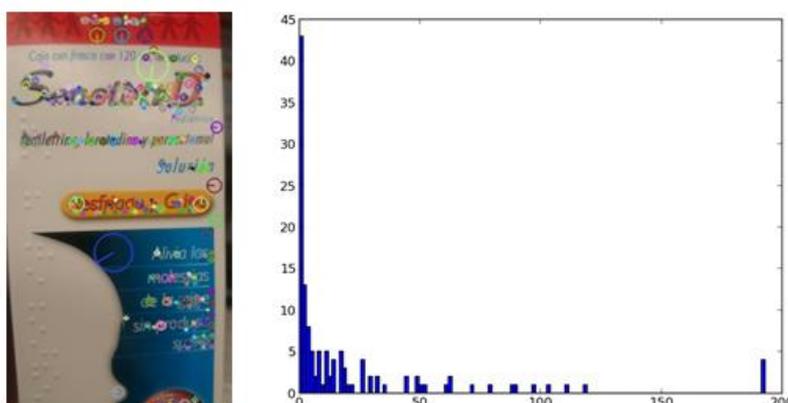
Fig. 3. A SIFT orientation histogram for one of the Sensibit D box keypoints on the left.

4.3   Dataset Creation

To create our dataset we applied the concept of the Bag-of-Words (BOW) method to the image's vectors extracted from the keypoints. The Bag-of-Words method is commonly used to classify text documents, where each text document can be represented by the frequency with which each word appears inside the document [27]. The process of the visual-word vectors' extraction is presented in the figure 4.

In order to use this method, the descriptor vectors extracted on the previous phase were grouped into clusters of similar descriptors. Each of these clusters represents a "visual word", in this manner an image can be represented as a Bag-of-visual-words (BOVW), then a visual vocabulary was created. Thus, for each image a vector is

generated containing the histogram of each visual-word. Then, these vectors can be used as features in a classifier [29-31].
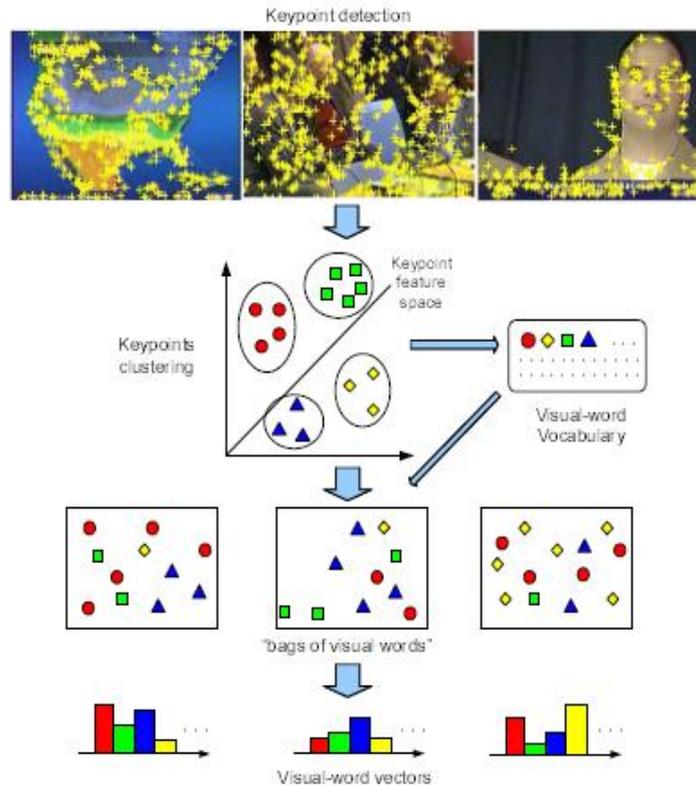
Fig. 4. Extracting the visual-word vectors [28].

For our project we used the K-means algorithm from the Scipy library to form clusters. This algorithm uses as parameter the number of clusters (k) to find. To determine the right k number of clusters we tried two of the methods presented in [32]. The first method we used was the rule of thumb in (2) for estimating the k number of clusters. This method gave us a value of $k \approx 143$.

$$k \approx \sqrt{n/2} \qquad (2)$$

The second method we used was the Elbow Method. This is a visual method to estimate the k number of clusters. The main idea of this method is to continuously increment the value of k until the cost function changes in a dramatic way, and then it reaches a plateau. However, sometimes the change cannot always be unambiguously identified. We used a range of k values from 50 through 450. As figure 5 shows using this method at the value of k=150 we observe the change.

Fig. 5. The elbow plot to choose *k*.

Fig. 6. Visual word histograms. At the upper row are histograms for Alka Seltzer boxes, and at the lower row are histograms for different Pepto Bismol bottles.

As both obtained values were close together, we decided to use their average as the number of k. Using $k \approx 147$, the K-means algorithm returned the centroids for each cluster, which means that we ended up with a vocabulary of 147 visual words.

Once we had this vocabulary, we extracted the frequency of visual words that exists for each image in order to obtain a histogram vector for each 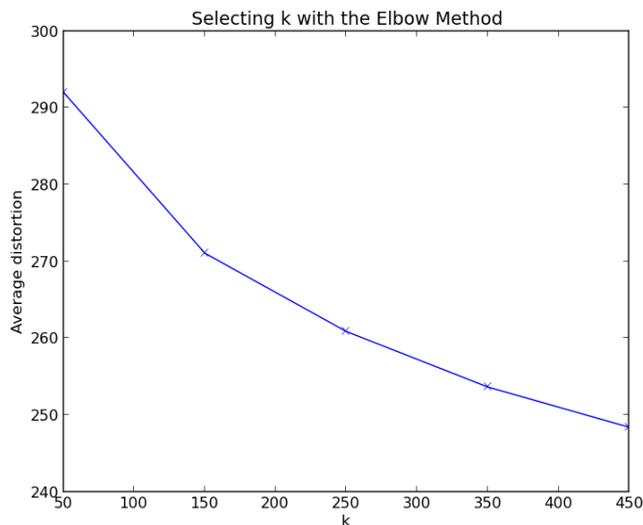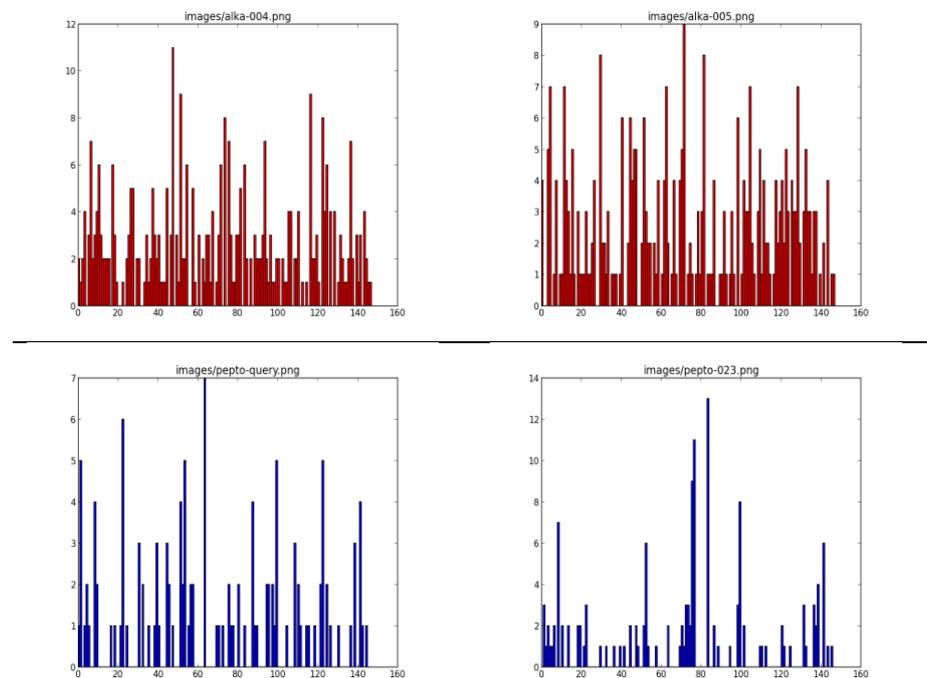of them. In figure 6 we can visualize the histogram for four different medication boxes. The upper row contains the histograms for two different Alka Seltzer boxes, and at the bottom row are the histograms for two different Pepto Bismol bottles. It can be seen the different histogram shapes for the two different counter medications and the similarity between the same medications.

4.4 Classifier Model Generation

We split the dataset in a train set with the 70% of the images (n=107), and the remaining images formed the test set (n=43). A stratified sampling was used with the intention that all the classes were equally distributed between the two datasets.

To generate the model we used Support Vector Machines (SVM) [33] algorithm. This is a supervised learning algorithm that has grown in popularity for classification and regression due to its excellent results [34, 35]. One of its advantages is the effectiveness when working with high-dimensional data and little number of samples [36]. Based on a labeled training data, this algorithm tries to find an optimal hyperplane that maximizes the separation margin between different classes. We used SVC implementation from the Scikit-Learn machine learning library in Python.

One important aspect to consider when using the SVM algorithm is to choose the right parameters to build the classifier. We test several values in a grid search strategy to obtain the best configuration parameters C and gamma, and also for the type of Kernel to use. The best parameters were for C=10, gamma=0.001, kernel='rbf'.

```
Training Confusion Matrix           Training Classification Report
[[ 7  0  0  0  0  1  0  0  0  0  0]               precision    recall  f1-score   support
 [ 0  6  0  0  0  0  0  0  0  0  0]
 [ 0  0 19  0  0  0  0  0  0  0  0]          0         1.00      0.88      0.93         8
 [ 0  0  0  8  0  0  0  0  0  0  0]          1         1.00      1.00      1.00         6
 [ 0  0  0  0 13  0  0  0  0  0  0]          2         1.00      1.00      1.00        19
 [ 0  0  0  0  0  8  0  0  0  0  0]          3         1.00      1.00      1.00         8
 [ 0  0  0  0  0  0  8  0  0  0  0]          4         1.00      1.00      1.00        13
 [ 0  0  0  0  0  0  1  7  0  0  0]          5         0.80      1.00      0.89         8
 [ 0  0  0  0  0  0  0  0 13  0  0]          6         0.89      1.00      0.94         8
 [ 0  0  0  0  0  0  0  0  0  8  0]          7         1.00      0.88      0.93         8
 [ 0  0  0  0  0  1  0  0  0  0  7]]         8         1.00      1.00      1.00        13
                                            9         1.00      1.00      1.00         8
                                           10         1.00      0.88      0.93         8

                                    avg / total       0.98      0.97      0.97       107
```

Fig. 7. Confusion matrix and classification report for the training dataset.

The SVM classifier for the training dataset had a 97.1% of accuracy, with an average of 98% for precision and 97% for recall. In figure 7 the confusion matrix for the training dataset is shown on the left, and the classification report is on the right. Precision is the ratio of how correct is our classifier when it says that an instance is positive. Similarly, the result of Recall is the proportion of positive cases that are classified correctly [36].

For its part, the F1-score is the harmonic mean that attempts to combine the two previous measures [36, 37]. In the same report it can be seen the result of the confusion matrix for the training data set. In this matrix, the diagonal elements represent the number of correctly classified elements; meanwhile those off-diagonal elements represent the number of incorrectly classified elements. The greater the number of elements found in the diagonal of the matrix, the better the classifier model.

5 Results

The resulting classifier was tested with the test dataset, which consisted on 43 images of counter medications. The accuracy for the classifier was of 90.7%, the confusion matrix for this test can be seen in figure 8. In addition, for this test the precision was of 93%, the recall of 91%, and the F1-score was of 91%.

```
Test Confusion Matrix          Test Classification Report
[[2 0 0 0 0 0 1 0 0 0 0]             precision  recall  f1-score  support
 [0 2 0 0 0 0 1 0 0 0 0]
 [0 0 8 0 0 0 0 0 0 0 0]        0      1.00     0.67      0.80       3
 [0 0 0 3 0 0 0 0 0 0 0]        1      1.00     0.67      0.80       3
 [0 0 1 0 5 0 0 0 0 0 0]        2      0.80     1.00      0.89       8
 [0 0 0 0 0 3 0 0 0 0 0]        3      1.00     1.00      1.00       3
 [0 0 0 0 0 0 3 0 0 0 0]        4      1.00     0.83      0.91       6
 [0 0 0 0 0 0 0 3 0 0 0]        5      1.00     1.00      1.00       3
 [0 0 0 0 0 0 0 0 5 0 0]        6      0.60     1.00      0.75       3
 [0 0 0 0 0 0 0 0 0 3 0]        7      1.00     1.00      1.00       3
 [0 0 1 0 0 0 0 0 0 0 2]]       8      1.00     1.00      1.00       5
                                9      1.00     1.00      1.00       3
                               10      1.00     0.67      0.80       3

                          avg / total  0.93     0.91      0.91      43
```

Fig. 8. Confusion matrix and classification report for the test dataset.

Fig. 9. New image classification and information retrieval.

To use the classifier model we developed a simple web application. From this application the user can upload the image of a counter medication and then ask the classifier to indicate the type of medication. Additionally, we created a database with

information about each medication. The result provided by the classifier is the input to query the database in order to retrieve the information related to the counter medication.

An example can be seen on the figure 9, where we used an image of the medicine Iliadin (Oximetazolina) taken directly from the smartphone's camera at a supermarket and uploaded to the web application though a web form. This image was not included on the initial dataset used to train or test the classifier. The uploaded image served as an input for the classifier and its output was used to query the database. As can be seen, the system predicted the correct medicine's class and returned the information regarding its name, compound, description and for what it is used. This information is returned in a JSON format to be used in other applications. The web application asks for the user feedback, in order to obtain more images for train the classifier seeking for accuracy improvement.

6   Conclusions

Because counter medication is easy to acquire in supermarkets, or convenience stores self-medication has become a common practice in many countries. Many studies in Mexico have shown that people use medication without prescription. In some cases this could have potential risks and consequences.

Machine learning and computer vision techniques are becoming a trend in healthcare applications in order to improve people's health. In this study we presented a classifier model aimed to help people who use counter medication to make better decisions. Our classifier uses SIFT and Visual-Bag-Of-Words techniques to extract features from counter medication boxes. With these features we generated a dataset that was used to train a SVM classifier. Our classifier got an accuracy of 90.7%, with a precision of 93%, a recall of 91%, and an F-1 score of 91% for the test dataset.

For testing purposes, the classifier was integrated in a web application to be used for people who wants to have more information about the counter medication he/she pretends to buy. In the same manner, this web application will help us to collect more images for training in order to improve the classifier's accuracy.

As a future work we want to expand the web application to include a recommendation system based on sentiment analysis, which uses our classifier to increase the information to be shown to users. This additional information will be extracted from Medic-Us, which is a social network for physicians and patients that allow the communication among them [38]. We also want to increment the number of counter medication classes known by the classifier, as well the number of images in dataset.

References

1. Peach, H.: Trends in self-prescribing and attitudes to self-medication. The Practitioner, 227(1384), pp. 1609 (1983)
2. Montastruc, J. L., Bagheri, H., Geraud, T., Lapeyre-Mestre, M.: Pharmacovigilance of self-medication. Therapie, 52(2), pp. 105–110 (1996)
3. Sanz, F., Loza, M. I., Ahlgrimm, E. D., Baetens, P., Cuypers, J., Cranz, H., Sosa-Iudicissa, M.: Telematics in community pharmacies to support responsible self-medication. The TESEMED project, Studies in health technology and informatics, pp. 141–148 (1996)
4. de Celis, E.S.P., Nava, Y. R.: Patrones de autoatención y automedicación entre la población estudiantil universitaria de la ciudad de Puebla. Elementos: ciencia y cultura, 11(56), pp. 43–51 (2004)
5. Ch, A.A.: Estudio de Automedicación en una farmacia comunitaria de la ciudad de Toluca. Revista Mexicana de Ciencias Farmacéuticas, pp. 5–11 (2009)
6. Ruiz, M.E.: Risks of self-medication practices. Current drug safety, 5(4), pp. 315–323 (2010)
7. Hughes, C.M., McElnay, J. C., Fleming, G. F.: Benefits and Risks of Self Medication. Drug safety, 24(14), pp. 1027–1037 (2001)
8. Alonso, A., Diaz, C., Sanz, F.: A matrix-approach proposal for the treatment of contents to be integrated into multimedia health resources for citizens. Studies in health technology and informatics, 68, pp. 3–6 (1998)
9. Acevedo, D.V., Alfaro Valle, A., Martínez Toledo, J. L.: Características de la adquisición de medicamentos en Morelia (Michoacán, México). Available from: http://iris.paho.org/xmlui/handle/123456789/15554 (1995)
10. Angeles-Chimal, P., Medina-Flores, M. L., Molina-Rodríguez, J. F.: Automedicación en población urbana de Cuernavaca, Morelos. Salud pública de México, 34(5), pp. 554–561 (1992)
11. Balbuena, F.R., Aranda, A. B., Figueras, A.: Self-Medication in Older Urban Mexicans. Drugs & Aging, 26(1), pp. 51–60 (2009)
12. Pagán, J.A., Ross, S., Yau, J., Polsky, D.: Self-medication and health insurance coverage in Mexico. Health policy, 75(2), pp. 170–177 (2006)
13. Patel, V.L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., Abu-Hanna, A.: The coming of age of artificial intelligence in medicine. Artificial intelligence in medicine, 46(1), pp. 5–17 (2009)
14. Szolovits, P.: Artificial Intelligence in Medicine. Boulder, CO, Westview Press (1982)
15. Buchanan, B.G., Moore, J. D., Forsythe, D. E., Carenini, G., Ohlsson, S., Banks, G.: An intelligent interactive system for delivering individualized information to patients. Artificial intelligence in medicine, 7(2), pp. 117–154 (1995)
16. Chu, K.C.: Applications of artificial intelligence to chemistry. Use of pattern recognition and cluster analysis to determine the pharmacological activity of some organic compounds. Analytical chemistry, 46(9), pp. 1181–1187 (1974)
17. Dasta, J.F.: Application of artificial intelligence to pharmacy and medicine. Hospital pharmacy, 27(4), pp. 312–315 (1992)

18. Poggio, T., Torre, V., Koch, C.: Computacional vision and regularization theory. Nature, 317(6035), pp. 314–319 (1985)
19. Gross, M.H.: Computer Graphics in Medicine: From Visualization to Surgery Simulation. ACM SIGGRAPH Computer Graphics, 32(1), pp. 53–56 (1998)
20. Van den Elsen, P.A., Pol, E. J., Viergever, M. A.: Medical image matching-a review with classification. Engineering in Medicine and Biology Magazine, pp. 26–39 (1993)
21. Ayache, N.: Medical computer vision, virtual reality and robotics. Image and Vision Computing, 13(4), pp. 295–313 (1995)
22. Pun, T., Gerig, G., Ratib, O.: Image analysis and computer vision in medicine. Computerized Medical Imaging and Graphics, 18(2), pp. 85–96 (1994)
23. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann (2005)
24. Kononenko, I.: Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine, 23(1), pp. 89–109 (2001)
25. Rahman, M.M., Bhattacharya, P., Desai, B. C.: A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance Feedback. IEEE transactions on Information Technology in Biomedicine, 11(1), pp. 58–69 (2007)
26. Ceh-Varela, E., Hernandez-Chan, G.: Tomatoes classifier using color histograms. In: International Conference on Advances in Engineering Science and Management (2015)
27. Layton, R., Learning Data Mining with Python. Packt Publishing Ltd. (2015)
28. Yang, J., Jiang, Y. G., Hauptmann, A. G., Ngo, C. W.: Evaluating bag-of-visual-words representations in scene classification. In: International workshop on multimedia information retrieval (2007)
29. Filliat, D.: A visual bag of words method for interactive qualitative localization and mapping. In: Proceedings IEEE International Conference on Robotics and Automation, IEEE (2007)
30. Yang, J., Jiang, Y. G., Hauptmann, A. G., Ngo, C. W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the international workshop on Workshop on multimedia information retrieval, ACM (2007)
31. Nicosevici, T., Garcia, R.: Automatic visual bag-of-words for online robot navigation and mapping. IEEE Transactions on Robotics, 28(4), pp. 886–898 (2012)
32. Kodinariya, T.M., Makwana, P. R.: Review on determining number of Cluster in K-Means Clustering. International Journal of Advance Research in Computer Science and Management Studies, 1(6), pp. 90–95 (2013)
33. Harrington, P.: Machine Learning in Action. Vol. 5, Manning Publications Co. (2012)
34. Cherkassky, V., Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), pp. 113–126 (2004)
35. Moreno, P.J., Ho, P. P., Vasconcelos, N.: A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In: Advances in neural information processing systems (2003)

36. Garreta, R., Moncecchi, G.: Learning scikit-learn: Machine Learning in Python. Packt Publishing Ltd. (2013)

37. Powers, D.M.: Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), pp. 37–6 (2011)

38. Hernández-Chan, G., Ceh-Varela, E., Jimenez-Kantun, M.: Medic-Us Red Social Inteligente. In: Avances y Perspectivas de la Innovación y Vinculación, Universidad Tecnológica Metropolitana, pp. 65–76 (2016)