International Journal on Emerging Technologies 11(2): 169-177(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Enhanced Feature Fusion and Selection Method for Improving Tablet Recognition

A. B. Dhivya[1] and M. Sundaresan[2]
[1]Research Scholar, Department of Information Technology, Bharathiar University, Coimbatore (Tamilnadu), India.
[2]Professor and Head, Department of Information Technology, Bharathiar University, Coimbatore (Tamilnadu), India.

(Corresponding author: A. B. Dhivya)
(Received 04 December 2019, Revised 30 January 2020, Accepted 05 February 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Acquiring knowledge regarding unidentified and/or mislaid tablets is beneficial to both physicians and individuals alike. Consumption of erroneous tablets results in adverse reaction and sometimes may even cause death. Thus, tablet identification that can provide accurate and fast information is vital. In this paper, an image query based tablet identification system is proposed, where the query image is the tablet image to be identified. This is performed using four steps, namely, preprocessing, feature extraction, feature selection and identification using the classifier. Features are attributes whose values make an occasion. The existing feature selection algorithm used ant colony optimization method to select relevant features in order to improve the performance of classification and identification. In proposed method, four sets of features namely, color, shape, statistical and texture, are extracted. In order to avoid the curse of dimensionality, a feature selection method that combined an enhanced ant colony optimization method with genetic algorithm is proposed. The enhancement of ant colony optimization is realized by combining it with an artificial bee colony. An SVM classifier is used during the identification of tablets. The accuracy and speed of SVM classifier is improved through the use of K-Means clustering and B+-tree indexing methods. Experimental results proved that the proposed method improves the performance of the identification system in terms of prediction accuracy and speed. The high accuracy and reduced search time obtained by the proposed system prove that the proposed tablet identification system can be safely used by both common public and healthcare professionals to identify unknown tablets.

Keywords: Ant Colony Optimization, Feature Fusion, Feature Selection, Genetic Algorithm, Indexing, Tablet Identification.

I. INTRODUCTION

The integration of information technology with healthcare industry has provided many sophisticated and efficient systems that help the physicians to identify diseases quickly at an early stage and increase quality of human life. Healthcare industries have increasingly amalgamated medical software with World Wide Web (WWW) to help doctors and patients in various tasks of diagnosis, treatment and post care. These software applications range from simple health tracking / fitness measuring to online evidence based medicine and intensive-care patient management. More often, the Internet is used as an encyclopedia of medical information resources and surveys have revealed that more than 80% of the WWW have used the Internet to obtain health information [1]. The WWW has become a potential environment to educate and empower the user by making available all those information on health and its related services and also provide self-help utilities. One such utility application is the tablet or pill identification system [2]. These systems when implemented in WWW are termed as e-Tablet Identification (e-TI) systems. The e-TI systems are defined as applications that use data mining techniques and/or image processing techniques to extract information or features regarding a tablet, which is then used to search a pre-established tablet database to retrieve similar tablet images. The extracted information can be, for example, external features (like name, disease symptoms, and manufacturer details) or appearance features (like color, size and shape details). These systems have tremendous uses in pharmacy, patient care & information systems, nursing [3] and aids in (i) decreasing inability during treatment and (ii) increasing patient confidence.

An e-TI system can work in two manners, namely, keyword-based and image-based system [4]. Keyword based e-TI system accepts as input, the details of a tablet like name, shape or color and performs a text-based search in tablet template database to retrieve matched tablets along with its information. On the other hand, image-based e-TI system accepts a tablet image as input and matches them with the images in the tablet template database and displays as output all similar images along with their respective information. As the only input required is the tablet image, the image-based e-TI systems, also called as Image Query-based Tablet Identification (IQTI) systems, are more accurate and fast. After accepting a query tablet image, the system performs various steps like preprocessing, feature vector creation, similarity computation and matching. The preprocessing of tablet images consists of tasks like image standardization and noise removal. The details regarding methods that dealt with enhancement procedures to improve the task of preprocessing has been explained [5,6]. It is a well-known fact that multiple features increase the accuracy of identification [7] and hence, the proposed IQTI system uses multiple features. One drawback while using multiple features is the curse of dimensionality. To solve this problem, a feature fusion and selection method is proposed to select only optimal features that can improve the accuracy of tablet identification. This method is termed as Enhanced Filter-Based Genetic Algorithm (EFGA) in this work. A hybrid ACO-ABC based feature selection algorithm combined with genetic algorithm (to produce fused feature set) is proposed in this work to reduce the curse of dimensionality problem. The selected optimal features are then used during matching step to retrieve similar images from the tablet template database. The proposed systems' matching performance is further improved through the use of a clustering based B+-tree indexing scheme. The following section is organized as follows. Section II presents the review of literature. Section III presents the methodology of the proposed IQTI system. Section IV presents the feature selection method. Section V explains the model construction. Section VI presents the results of the experiments conducted to evaluate the proposed methods. Section VII concludes the work and Section VIII ends with future research directions.

II. REVIEW OF LITERATURE

Ladha & Deepa (2011) proposed a way to evaluate the feature selection method which shows the behavior of relevance, irrelevance, redundancy and sample size of synthetic data sets. A set of experiments using different datasets are carried out and provides a set of optimal solutions. These sets of solutions are compared with the output of feature selection methods and their degree of approximation to the real solution was measured using the scoring measure [8].

Karaboga and Ozturk (2012) describes the feature selection method that makes it possible to obtain more accurate results by removing irrelevant and disconnected features in model prediction. The model calculation provides the efficient relationship between the output attribute y and the input attribute x of the data set. Removing unrelated features minimize the aspect of the model, thus it reduces space complexity and computation time [9].

Imani *et al.*, (2017) worked on with ACO method and Genetic algorithms using support vector machine as a classifier for Persian font's database classification and lead a better accuracy compared to other methods [10].

Prasartvit *et al.*, (2013) discussed in their work that how the Artificial Bee Colony method is used for selecting features from the high dimensional data. It is used for selecting the best and relevant features using the ABC. This work uses the wrapper based approach for the feature subset evaluation that's why the K-nearest neighbor classification technique is used along with the ABC method [11].

Dwivedi *et al.*, (2019) proposed a method of feature selection algorithm based on Ant Colony Optimization (ACO) is that removes the unimportant, irrelevant and redundant features and selects the more appropriate features from data having large number of features. This method uses one of the swarm intelligent algorithm for feature selection based on ant colony optimization. This algorithm is combined with the Support vector machine classifier for selecting the more appropriate and useful features [29].

III. METHODOLOGY

The hunt for techniques that can improve the process of searching and retrieving similar tablet images from huge databases is the most important and challenging job in IQTI systems. This involves two main steps, namely, feature extraction and matching. In order to improve the performance of the proposed IQTI system, two more steps, namely, Feature Selection and clustering-based matching and indexing are included. The methodology used by the proposed IQTI system is shown in Fig. 1 and all the steps involved are enhanced in order to improve the overall performance of IQTI system.

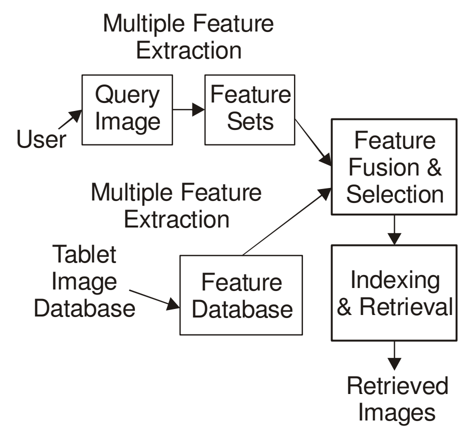

Fig. 1. Steps in the Proposed IQTI System.

Feature Extraction: Feature extraction is defined as the construction of a compact feature vector that can efficiently represent the interesting or important regions of a tablet image. During the process of identifying a tablet, it is always beneficial to have details regarding various parameters of a tablet, like color description, shape description, etc. instead of using a single descriptor. Accordingly, in this work multiple features are extracted from a tablet image. The features extracted are grouped into four categories as given in Table 1.

Table 1: Features Extracted for Text Recognition.

Category	Features Extracted
Color Features (CF)	Color Layout Descriptor (CLD), Dominant Color Descriptor (DCD)
Shape Features (SF)	Edge Histogram Descriptor (EHD), Pyramid Histogram Descriptor (PHD), Pseudo-Zernike Moment (PZM), Polar Harmonic Transform (PHT)
Statistical Features (STF)	Area (A), Perimeter (P), Compactness (C), Eccentricity (EC), Elongation (EL), Ratio (R) and Magnitude (M)
Texture Features (TF)	Gabor Filter (GF), Directional Local Motif XoR Pattern (DLMXoRP)

These multiple sets of features, while increasing the accuracy, also increase the size of the feature vector constructed which in turn, increases the computational complexity of the IQTI system. To solve this issue, a method that constructs a fused optimal feature vector is proposed in this work.

The main aim of this method is to create a feature vector that has multiple feature characteristics and implements the three requirements of optimality, namely, minimum redundancy, maximum relevancy and maximum discriminating capacity. The working of this method is described in the following section.

IV. FEATURE SELECTION USING ENHANCED FILTER-BASED GENETIC ALGORITHM

Feature Selection (FS) is a process that is used to reduce the size of the feature vector by selecting only relevant features, which increase the accuracy and removing all redundant and unwanted features [12]. In other words, it is the process of finding most relevant inputs to the pill identification system. The FS method belongs to any one of the three main categories, namely, filter-based, wrapper-based and embedded approaches [13]. Recent years have envisaged several heuristic methods for FS in different applications [14, 15], among which the Genetic Algorithm (GA) is more popular [16].

A GA can solve problems using a natural selection procedure that mimics biological evolution, to iteratively modify a population of individual solutions to find an optimal solution [17].

At each step, from the current population, it selects random individuals called parents, which are used to produce children for next iteration or generation. More than consecutive iterations, the population evolves to a best possible solution. In this work, the solution is an optimal feature set that can enhance the process of pill identification.

Usage of GA for FS, apart from working better than conventional approaches, also provide many advantages including its ability to work with feature sets that has multiple features and not requiring knowledge regarding the problem under study. However, it also suffers from disadvantages like (i) being computationally expensive because the evaluation of each individual is required during the building of the identification model and (ii) long convergence time because it has a stochastic nature. Both of which can be solved by avoiding irrelevant regions in the search space. Avoiding irrelevant features can be achieved in two manners. The first is in the generation of initial solutions and the second is to reduce the feature set size searched by GA.

In this work, both these steps are incorporated. The initial solutions are generated using Ant Colony Optimization (ACO), which also acts as a method to obtain an initial set of optimal features. The initial solutions obtained are embedded in the initial solutions of the enhanced GA. The reduction in feature size is obtained through the use of a classification-based feature selection method.

The proposed FS method consists of two steps. The first is the reduction step and the second is the optimal feature selection and fusion step. The reduction step, whose goal, as the name implies, is to generate a reduced version of the multiple feature sets extracted in feature extraction. For this purpose, an enhanced Ant Colony Optimization (ACO) method is proposed.

The goal of the second step (fusion step) is to combine the multiple reduced feature sets from step 1 into a single feature vector. In the second step, the GA is used to obtain the fused, optimal feature set. The output from the application of step 1 produces the restricted feature space, which can reduce the execution time of the GA. The flow of EFGA is figuratively shown in Fig. 2.

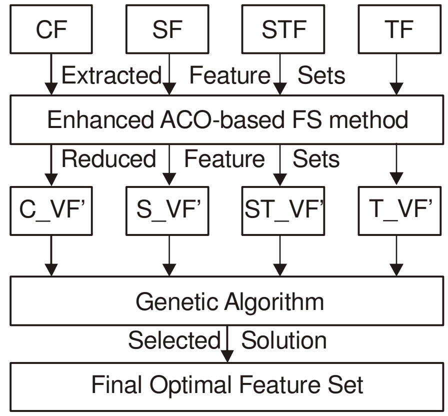

Fig. 2. Steps in EFGA.

Reduction Step: As mentioned in the previous section, four sets of features, namely, CF, SF, STF and TF, are extracted from the tablet images. The first step of EFGA uses an enhanced ACO-based feature selection method. An ACO method [18] is a branch of swarm intelligence which uses collective intelligent groups of simple agents to optimize the task of feature selection. It is a discipline that deals with the natural and artificial systems composed of a group of insects (examples include schools of fish, colony of ants, bees and flock of birds), where each individual insect can perform a simple task on its own and the colony's cooperative work is the main reason determining the intelligent behavior it shows [19, 20]. In simple terms, swarm intelligence methods are stochastic search methods that mimic the natural biological evolution and/or the social behavior of the species [21]. Among the various swarm intelligence methods, the ACO method is inspired by the social behavior of the ants. Ants, as such, do not have sight, but have the capability to search the shortest route from their nest to food source using chemical materials called pheromone, which is left behind when they move. This pheromone can be smelled by other ants and which gives them the ability to follow the same path that has been successively passed.

The concept of using ACO was first conceived to solve the travelling salesman problem [22], after which has been successfully used in a large number of applications. The motivation behind using this method for feature selection lies in the act that there seems to be no heuristics that can guide search to the minimal optimal feature subset every time. When a graph is used to represent the features, the ants can discover the best combination of features just by traversing the graph. However, this faces the issue of requiring prior knowledge of the features. In this work, this issue is solved, thus enhancing the ACO based FS. The ACO method when compared to other existing protocols that provides maximum advantage during feature selection. However, it has two other main issues, namely, slow convergence and tendency to stagnancy. Both these issues are solved by combining ACO with ABC.

An ACO method, while used for feature selection, is required to be represented in a graph, whose nodes denote the feature and the edges between the nodes representing the choice of next features. The optimal feature subset search is then described as a task where an ant move through the graph in a direction that requires a minimum number of node visits that satisfies a stopping criterion. Let k be an ant and i and j be two features. Let F be the original full feature set and F' be the final optimal feature subset. Eqn. (1) represents the probabilistic transition rule, which describes the probability of k at feature i choosing j at time t.

$$P_{ij}^{k}(t)=\begin{cases}\dfrac{(\tau_{ij}(t))^{\alpha}(\eta_{ij})^{\beta}}{\sum_{l\in J_i^k}(\tau_{il}(t))^{\alpha}(\eta_{il})^{\beta}} & j\in J_i^k\\ 0 & \text{Otherwise}\end{cases} \quad (1)$$

Here, η_{ij} represents the recognition of an ant (at node i) to pick feature j because the next node to visit. J_i^k is the set of neighboring nodes at node i that have not been visited yet by k. The two parameters, $\alpha > 0$ and $\beta > 0$ are taken to approximate the virtual consequence of the pheromone value and therefore the heuristic information. The values of α and β were determined experimentally and was set as 1 and 0.1 respectively, after empirical estimation. In the equation, $\tau_{ij}(t)$ represents the quantity of pheromone on edge (i, j).

In the present work, instead of using the above described conventional manner, a bee is created and is sent to the selected nodes. The fitness function used is the classification accuracy. The node with the highest pheromone, which satisfies both Eqn. (1) and bee fitness function is taken as the best bee and is added to the feature set solution.

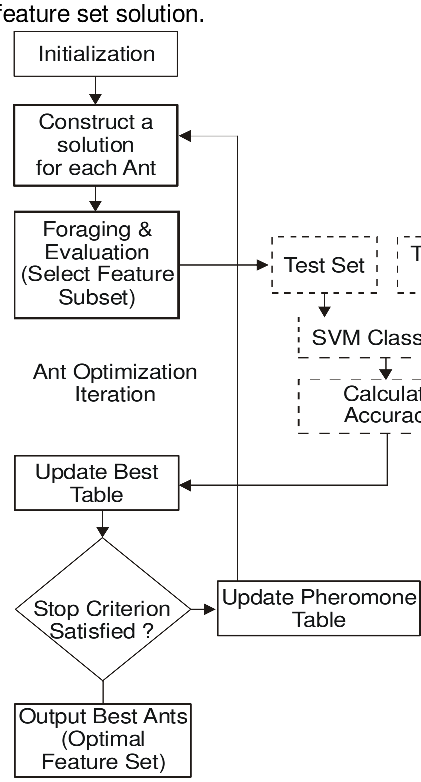

Fig. 3. ACO Based Feature Selection.

The steps involved during ACO-based feature selection are given in Fig. 3, which performs feature selection using two tasks. The first is the ant colony optimization task (denoted by straight boxes) and the second is the performance evaluation using SVM classifier (denoted by dashed boxes). The method begins with ACO initialization step, which generates a set of ants (say, m) and set them randomly on the graph. This means that the each of the m ants start with one feature selected randomly. In this work, the number of ants, that is, m is set to the number of attributes or features extracted. This enables each ant to start the construction of path at different feature. Thus, in this work, each node represents a feature and the edge between two nodes is the paths between the nodes. During initialization, the initial intensity of pheromone trails associated with any feature is also set to one. In this step, the maximum number of allowed iterations, T, is set to 200. Thus, the stopping or convergence criterion is 'if Number of iterations is equal to T, then stop'. At this junction, the proposed method outputs the best feature set as optimal set.

After initialization, the next step is focused on finding the next probable node to visit. At this point, the ants are replaced by bees, which begin to traverse the edges probabilistically. Each bee should visit all features during a traversal and construct solutions completely. The constructed solutions then are fed to the second task, that is, performance evaluation task. This task, initially trains a Support Vector Machine (SVM) classifier using the training set of the original feature set. This trained model is then used to test the feature set solutions produced by the ACO task and the accuracy of classification estimated. If a solution is not able to increase the classification accuracy for two consecutive steps, the method completes its work and exits.

All the resulting subsets are then evaluated and the one with optimal performance is considered as the best or optimal feature subset. During the best feature set selection, the evaluated subsets are arranged in descending order of their accuracy and the first set is selected (this is the set with maximum accuracy) and which exceeds an accuracy threshold. If this condition is not reached, the pheromone is updated, a new set of ants is generated by removing the previous ants and generating new units. The whole process is then repeated with this new set, until either the stopping criterion is met or till an optimal feature set is obtained. The pheromone on each edge is updated using Eqn. (2).

$$\tau_{ij}(t+1)=(1-\rho).\tau_{ij}(t)+\sum_{k=1}^{m}\Delta_{ij}^{k}(t) \quad (2)$$

Where, $\Delta_{ij}^{k}(t)=\begin{cases}\gamma(S^{j})/|S^{k}| & \text{if } (i,j)\in S^{k}\\ 0 & \text{Otherwise}\end{cases}$.

Here, ρ, which can take up a worth between 0 and 1, could also be a decay constant and is used to simulate the evaporation of pheromone. In the above equation, S^k is the feature subset found by ant k. The updating of the pheromone takes into consideration both the goodness measure (γ) of the ant's feature subset and the subset size. This method allows all ants to update the pheromone.

Feature Fusion and Selection: The goal of this step is to absorb the optimal results from multiple feature sets, obtained through the feature selection method, in order to find a combined feature set, which has an informative feature that has the maximum discriminating capacity and which can improve the performance of the pill identification system. In this work, the Genetic Algorithm (GA) is used to realize this goal. The set of feature sets obtained from the previous step is used to form a feature pool. This feature pool will have a collection of candidate features, from which the GA will find a set of optimal feature subset, which will exhibit characteristics of multiple features. The steps involved during GA-based fusion and selection is presented in Fig. 4. The GA Based Feature Fusion and Selection is used to handle the curse of dimensionality.

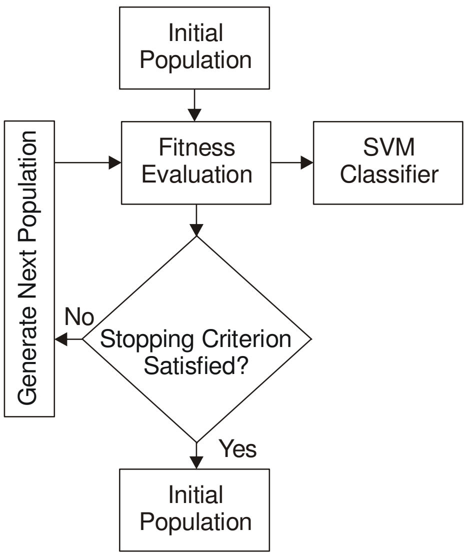

Fig. 4. GA Based Feature Fusion and Selection.

The first step, initialization takes care of the representation of hypotheses, where each feature subset is an individual in a population and is encoded (or represented) as an n-bit binary vector. A bit with value 1 indicates that the corresponding feature is selected and value 0 means that the feature is not selected. The fitness function used is given in Eqn. (3).

$$F = w * c(x) + (1 - w) * (1/s(x)) \tag{3}$$

In the above equation, x is the feature subset selected, c(x) denotes the classification accuracy, s(x) denotes the size of x. Any feature selection method works towards increasing the classification accuracy while minimizing the optimal feature set size. To achieve this tradeoff, a weight factor w, associated with the classification accuracy and feature subset size, is used. The fitness of x will (i) increase with increase in classification accuracy and (ii) decrease with increase in size of x. When a higher value of w is used, it indicates that the classification accuracy gives more priority. Alternatively, a low w value indicates that more penalties are given on the size of x. Thus, achieving a mid-point, a tradeoff between accuracy and optimized feature set size can be obtained. The value of w is always between 0 and 1, and in this work it is set to 0.75 after empirical evaluation.

In this work, the induction method used is the Support Vector Machine (SVM) classifier.

The proposed GA based feature selection and fusion method use the three genetic operators, selection, crossover and mutation, for generating a new population. This work uses the Roulette wheel selection operator, which selects an individual feature subset probabilistically from a population for latter breeding. Eqn. (4) presents the probability of selecting an individual n_i.

$$P(n_i) = \frac{F(n_i)}{\sum_{i=1}^{p} F(n_i)} \tag{4}$$

where $F(n_i)$ is the fitness value of n_i. The individual n_i will be selected, if the resulting probability is directly proportional to its own fitness and inversely proportional to the fitness of other competing hypothesis in the current population.

Mutation and crossover are two of the foremost commonly used operators with genetic algorithm that represent individuals as binary strings. Mutation operates on one string and usually changes a touch randomly. Crossover operates on two parent strings to supply two off springs. In this work, the crossover operator used is the single-point operator, where a point is randomly selected so that the first i bits are obtained from one parent, while the remaining bits are contributed by another or second parent. During feature fusion and selection, each individual has a probability to mutate. In this work, n bits to be flipped are randomly selected during each mutation stage.

V. MODEL CONSTRUCTION AND IMAGE RETRIEVAL

Identification of techniques to improve the process of searching and retrieving the interested images from large databases is a challenging task in IQTI systems. This involves various steps like extraction of content descriptors (features), matching along with indexing schemes to improve the retrieval process.

The increasing methods to create digital images indicates that the conventional methods for searching images alone is not sufficient to perform effective and efficient retrieval and demands for new sophisticated methods to be identified and developed. One major issue during query phase (after feature extraction) is that the huge number of computations and comparisons required to seek out the relevant images. In this paper, a classification and clustering based scheme combined with B+-tree indexing method is used to reduce the number of distance (similarity) calculations.

The proposed matching and retrieval uses a sequence of steps to retrieve relevant tablets that match the query tablet images and are given in Fig. 5. The first step is clustering, which is done using K-Means Algorithm. After clustering, details regarding the tablet images in each cluster and its centroid are calculated. These values are then used to construct one B+-tree per cluster. The indexing scheme of the B+ - tree method reduces the total I/O cost that eradicates the major bottleneck accompanied with large image databases.

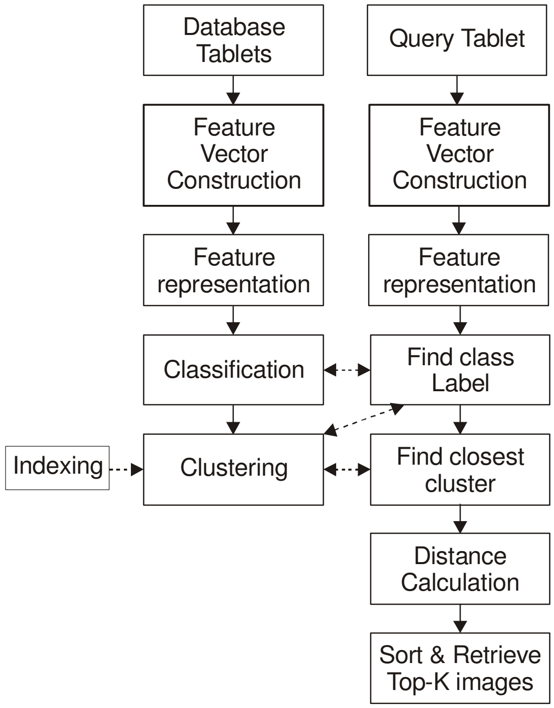

Fig. 5. Model Construction and Image Retrieval.

Further, dynamically changing the indexing data structure of the B+ - tree makes the proposed system extensible. That is, if a user wants to add new images to the database, the B+-tree can handle it efficiently with minimum effort.

An important step during image retrieval is to find a good similarity measure that has the best ability to find images. The similarity measures determine how the feature vectors are distributed in the feature space, affecting critically how the vectors correspond to perceptual qualities. In general, similarity measures are used to find the top-k images that have a minimum distance between feature points of database and query images [23]. As rightly pointed out by the author [24], effectiveness by comparing two images heavily relies on the definition of a good similarity measurement. Finding the similarity between the query and database image is a subjective decision, where the use of class labels of the images, to provide valuable information, is advantageous. In this paper, a similarity measure that combines two frequently used methods is proposed to further increase the accuracy of matching and retrieving. The two similarity measures used in this work are Euclidean distance and Manhattan distance. The method of combining these two similarity measures is presented in Fig. 6.

The query step performs the same processes of feature extraction and reduction. Using this reduced feature vector, it is classified to the closest label using the classification method. A distance function is then used to calculate the distance between the query image feature vector and the representative feature vectors of all clusters having the same class of images as the query image. The cluster with the minimum distance from the query image is then looked for matches. Top-K images from the matched group are returned as retrieval result.

Let TQ denote the query tablet image and TD denote the tablet images in the database
Let F_i be the optimal feature set
Using Euclidean distance between TQ(F_i) and TD(F_i)
Order them according to distance
Retrieve top 70% and store them separately as most relevant tablets (RT)
Now use Manhattan distance to calculate distance between TQ(F_i) and RT(F_i)
Order them according to distance
Retrieve top 10 images as selected similar images

Fig. 6. Multiple Similarity Measure.

After clustering, details regarding the images in each cluster and its centroid are known. For each image, the distance between the image and the cluster center that the image belongs to is calculated using similarity measures. In the query phase, the query image's distance to cluster centroid and the queried image's distance to cluster centroid is used as a similarity measure. The distance measure between querying image and its centroid is stored in the model, so as to reduce distance calculations during the query phase. This will also reduce the search space. These values are then used to construct one B+-tree [25, 26] per cluster. Fig. 7 presents the steps during model construction.

```
Call K-Means Algorithm
//Output will be cluster centers and the cluster of
each image in the database
for all cluster i do
        Initialize a B+-tree, BT
        c = cluster center i
        for all image j in cluster i do
            d = distance between cluster center i
and feature vector of image j
            Add (d, image name) to BT where d is
the key
        end for
end for
Return all BTs
```

Fig. 7. Model Construction Algorithm.

VI. EXPERIMENTAL RESULTS

Several experiments were conducted to evaluate the performance of the proposed methods to study their effect on tablet identification. The image database used during evaluation consists of tablet images stored in JPEG format. The images belonged to two categories, namely, reference and consumer quality images. Reference quality images are high quality images. This set of images has both their frontal and back view stored and was downloaded from ftp://nlmpubs.nlm.nih.gov/nlmdata/pir/dr.zip [27]. Consumer quality images are images captured using a digital camera or mobile phones, as normal public will do and is available at http://www.ftpdir.hu/nlmpubs.nlm.nih.gov/nlmdata/pir/ [28]. Totally ten images from the database were used for the experimental purpose. Examples of reference and consumer images are shown in Fig. 8. Five performance metrics, namely, precision, recall, f-measure, accuracy and speed were used to evaluate the effect of the proposed methods on the performance of the IQTI system. The abbreviation used during discussion is given in Table 2.

Fig. 9 to 14 show the effect of the feature selection methods on tablet identification and the efficiency of the various IQTI systems in terms of the precision, recall, F-measure, accuracy and speed respectively.

From the experimental results, it is clear that, use of feature selection methods to identify tablets is the beneficiary. This is evident from the high values obtained by the various methods while using feature selection when compared to 'No FS'. Maximum efficiency was produced by the proposed EFGA method.

Reference Images	Consumer Images

Fig. 8. Test Images of Tablets.

Table 2: Abbreviation Used.

Abbreviation	Description
NO FS	No Feature Selection
ACO	Feature Selection Using ACO Method
ACO-ABC	Feature Selection Using ACO Combined with ABC Method
GA	Feature Selection Using GA
EFGA	Enhanced Filter-Based Genetic Algorithm (EFGA) using ACO and ABC
IQTI-SVM	IQTI System using SVM Classifier
IQTI-SK	IQTI System using SVM Classifier and K-Means Clustering
IQTI-SB	IQTI System using SVM Classifier with B+-Tree Indexing
IQTI-SKB	IQTI System Using SVM Classifier, K-Means Clustering and B+-Tree Indexing

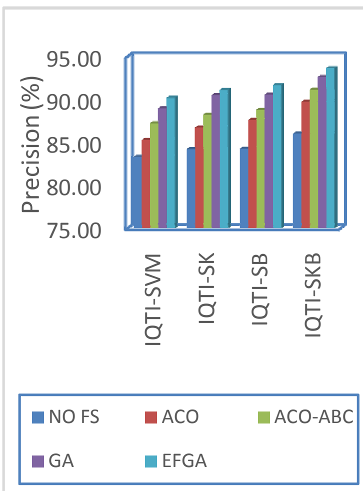

Fig. 9. Precision (%).

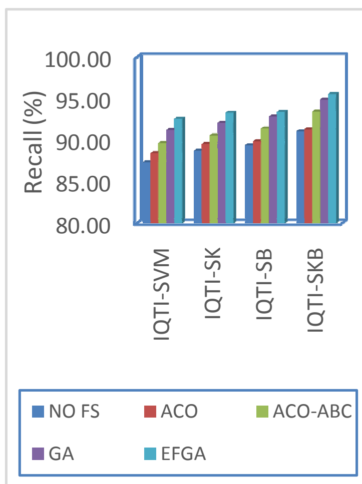

Fig. 10. Recall (%).

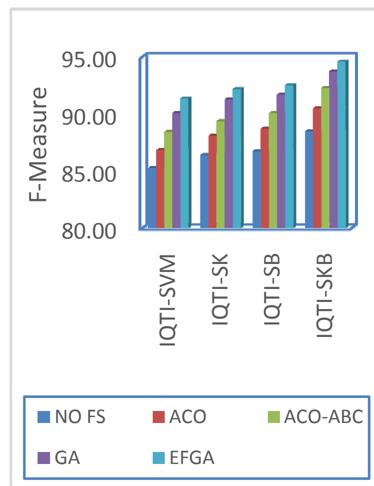

Fig. 11. F-Measure.

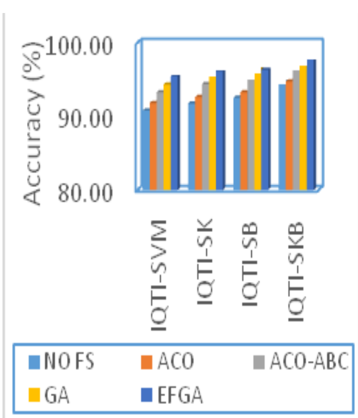

Fig. 12. Accuracy (%).

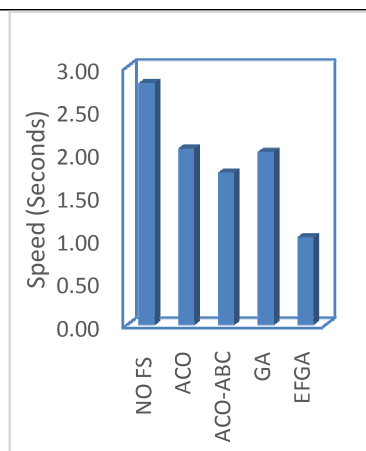

Fig. 13. Speed (Seconds) of FS.

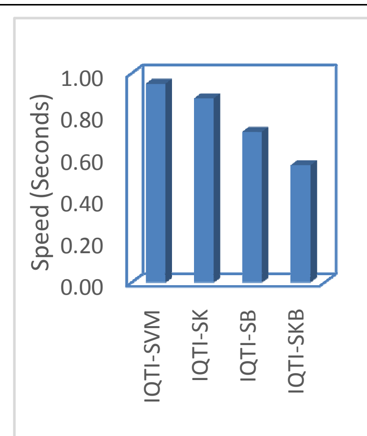

Fig. 14. Speed (Seconds) of Tablet Identification Systems.

Comparison between IQTI systems showed that the results further prove that the use of the indexing method improves the execution time and works even better than the usage of clustering methods. The proposed system that used SVM classifier, K-Means clustering and B+-tree indexing methods proved to be highly efficient while using the optimal feature subset produced by EFGA method. This method was efficient by 7.17% when compared to conventional SVM based IQTI system with no feature selection. Thus, it can be concluded that the proposed feature selection method has a positive impact on tablet identification and the method of combining classification, clustering and indexing is highly successful.

VII. CONCLUSION

This paper proposed enhanced methods for the two steps in IQTI, namely, selecting optimal features and identifying tablets. Four types of features were extracted in this work. They are color features, shape features, statistical features and texture features. In order to solve the problem of high dimensionality, in this paper, an enhanced feature selection method that is an amalgamation of three techniques, namely, Ant Colony Optimization, Artificial Bee Colony and Genetic algorithm was proposed. Further, to improve the process of searching tablet images that are similar to the query image, two methods, clustering and indexing was combined with classification. The classifier, clustering and indexing methods used were support vector machine, K-Means and B+-tree methods. The experiments evaluating the proposed method proved that the features selected by the proposed method were most informative and reduced the size of feature vector used. This, in turn, reduced the execution time of tablet identification. Further, reduction in feature set size enabled the classifier to learn a more robust solution and thus achieved a better identification performance. The usage of clustering and indexing also reduced the searching time, while increasing the identification accuracy. Thus, it can be concluded that the proposed IQTI system can be safely used by both physicians and individuals to retrieve similar tablets efficiently.

VIII. FUTURE SCOPE

Future researches are planned to consider other clustering and indexing method to combine with classification to further improve tablet detection.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Head of the Department and entire staff of the Department of Information Technology, Bharathiar University, Coimbatore.

Conflict of Interest. The Author(s) declare(s) that there is no conflict of interest.

REFERENCES

[1]. Dang, L.M., Piran, J., Han, D., Min, K., & Moon, H. (2019). A survey on Internet of things and cloud computing for healthcare, Electronics, MDPI., Article ID, 768, *8*(7): 1-49.

[2]. Faggella, D., (2018). Machine learning healthcare applications - and Beyond, Techemergence, https://www.techemergence.com/machine-learning-healthcare-applications.

[3]. IEEE Technical Committee (2019), Automation in Health Care Management, IEEE Robotics & Automation Society, http://www.ieee-ras.org/automation-in-health-care-management.

[4]. Chen, R.C., Chan, Y.K., Chen, Y. H., & Bau, C. T. (2012). An automatic drug image identification system based on multiple image features and dynamic weights, *International Journal of Innovative Computing, Information and Control.*, *8*(5), 2995-3013.

[5]. Dhivya, A. B. & Sundaresan, M. (2016). Performance analysis of interpolation methods for improving sub-image content-based retrieval, Proceedings of the INDIACom, *3rd International Conference on Computing for Sustainable Global Development*, 1-5.

[6]. Dhivya, A. B. & Sundaresan, M. (2019). Enhancing Tablet Images Using Denoising Algorithms in Various Color Space Models, *Future India: Science & Technology, Proc. 106th Indian Science Congress, Jalandhar (Abstracts of Best Poster Award Programme).*, 3-7, 24.

[7]. Li, R., Lu, W., Liang, H., Mao, Y. & Wang, X. (2018). Multiple Features with Extreme Learning Machines for Clothing Image Recognition, *IEEE Access, 6*, 36283-36294.

[8]. Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. *International journal on computer science and engineering*, *3*(5), 1787-1797.

[9]. Karaboga, D., & Ozturk, C. (2012). A novel clustering approach: artificial bee colony (ABC) algorithm, *Applied Soft Computing Journal*, *11*(1), 652–657.

[10]. Imani, M. B., Pourhabibi, T., Keyvanpour, M. R., & Azmi, R. (2017). A new feature selection method based on ant colony and genetic algorithm on Persian font recognition, *International Journal of Machine Learning and Computing*, 278-282.

[11]. Prasartvit, T., Banharnsakun, A., Kaewkamnerdpong, B., & Achalakul, T. (2013). Reducing bioinformatics data dimension with ABC-kNN. *Neurocomputing*, *116*, 367-381.

[12]. Wang, Y., Feng, L. & Zhu, J. (2018). Novel artificial bee colony based feature selection method for filtering redundant information, *Applied Intelligence*, *48*(4), 868-885.

[13]. Gonzalex, S., Herrero, A., & Sedano, J. (2019). Neuro-Evoluitonary Feature Selection to detect android malware, Advances in Intelligent Systems and Computing, Alvarez, F. M., Lora, A.T., Munoz, J.A.S., Quintian, H., Corchado, E. (Eds.), *Proceedings of 12th International Joint Springer Conference on Computational Intelligence in Security for Information Systems and 10th International Conference on European Transnational Education*, 1-17.

[14]. Al-Harbi, O. (2019). A Comparative Study of Feature Selection Methods for Dialectal Arabic Sentiment Classification Using Support Vector Machine, *International Journal of Computer Science and Network Security*, *19*(1), 167-176.

[15]. Sobhanzadeh, Y. M., Motieghader, H., & Nejad, A. M. (2019). Feature Select: a software for feature selection based on machine learning approaches, *BMC Bioinformatics.*, Article ID 170, *20*, 1-17.
[16]. Wutzl, B., Leibnitz, K., Rattay, F., Kronbichler, M., Murata, M., & Golaszewski, S. M. (2019). Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness, *Plus One*, Article ID e0219683, *1*(7), 1-16.
[17]. McCall, J. (2005). Genetic algorithms for modelling and optimization, *Journal of Computational and Applied Mathematics*, *184*(1): 205-222.
[18]. Dorigo, M., & Di Caro, G. (1999). Ant colony optimization: a new meta-heuristic. In *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, *2*, 1470-1477.
[19]. Liu, B., Abbass, H. A., & McKay, B. (2004). Classification rule discovery with Ant Colony Optimization, *IEEE Computational Intelligence, 3*(1), 31-35.
[20]. *http://www.scholarpedia.org/article/Swarm_intelligence*., Last Accessed during December, 2019.
[21]. Elbeltagi, E. E. (2013). Swarm Intelligence for Large-Scale Optimization in Construction Management, *Metaheuristic Applications in Structures and Infrastructures, Elsevier*, 479-495.
[22]. Hussain, A., Muhammad, Y. S., Sajid, M. N., Hussain, I., Shoukry, A. M., & Gani, S. (2017). Genetic Algorithm for Traveling Salesman Problem with Modified Cycle Crossover Operator, *Computational Intelligence and Neuroscience*, Article ID 7430125, 1-7.
[23]. Bi, J., Chen, Y., & Wang, J. Z, (2005). A sparse support vector machine approach to region-based image categorization, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR "04)*, *1*, 1121-1128.
[24]. Moosmann, F., Nowak, E., & Jurie, F. (2008). Randomized clustering forests for image classification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(9), 1632–1646.
[25]. Lightstone, S., Teorey, T., & Nadeau, T. (2007). Physical Database Design: The Database Professional's Guide to Exploiting Indexes, Views, Storage, and More, *Morgan Kaufmann Publishers*.
[26]. Powell, G. (2006). Beginning Database Design and Implementation, *Wrox Press*.
[27]. ftp://nlmpubs.nlm.nih.gov/nlmdata/pir/dr.zip Last Accessed during December, 2019.
[28]. http://www.ftpdir.hu/nlmpubs.nlm.nih.gov/nlmdata/pir/La st Accessed during December, 2019.
[29]. Dwivedi, R., Kumar, R., Jangam, E., & Kumar, V. (2019). An Ant Colony Optimization Based Feature Selection for Data Classification. *International Journal of Recent Technology and Engineering (IJRTE), 7*, 2277-3878.

How to cite this article: Dhivya, A. B. and Sundaresan, M. (2020). Enhanced Feature Fusion and Selection Method for Improving Tablet Recognition. *International Journal on Emerging Technologies*, *11*(2): 169–177.