Available online at: http://www.ijmtst.com/vol6issue03.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 06, Issue No: 03, March 2020

A New Abnormal Red Blood Cell Data Set

Sherna Aziz Toma[1] | RajaaSalih Mohammed Hasan[2]| Loay E. George [3]| Azizah Bet. Suliman[4]

[1]College of medicine, Baghdad University, Baghdad, Iraq.

[2]Al Mansur Institute of Medical Technology. Middle Technical University Baghdad, Iraq

[3] University of Information Technology and Communication. Baghdad, Iraq

[4] College of Information Technology, Tenaga National, Malaysia

To Cite this Article

Sherna Aziz Toma, RajaaSalih Mohammed Hasan, Loay E. George and Azizah Bet. Suliman, "A New Abnormal Red Blood Cell Data Set", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 03, March 2020, pp.:93-97.

Article Info

Received on 19-February-2020, Revised on 27-February-2020, Accepted on 05-March-2020, Published on 11-March-2020.

ABSTRACT

Blood diseases a worldwide public health problem such as thalassemia, anemia malaria. It diagnosis basis in abnormal shape,size and color of red blood cells (RBC). in this paper the data were obtained from 1000 RBCs that extract from 150 blood smear images for different diseases. After that, several image pre-processing steps done to extract statistical features using a new system. k-means well-clustering algorithm was used to solve the challenges of the color variability of both the background and the cell's pixels is the main problems we met in this work. 100 features for 30 types of normal and abnormal RBCs. The anew data set was test with different machine learning classification algorithms. The result show very high accuracy more than 93%, and evaluation algorithms models achieve more than 94.4% These result conclude that the algorithms have high ability of distinguish between the true negative and true positive more than 95% and Root mean squared error lease than 0.38, the fact that the features are very accurate in describing RBCs.

KEYWORDS: *abnormal red blood cells; image processing;k-means, feature extraction, Machen learning algorithms*

I. INTRODUCTION

Normal cell morphology is the key to the maintenance of cellular functions. Any changes of cell structure will impair or cause loss of cell function (Deyi,2012). Anemia is one of the most important diseases whose diagnosis depends on the shape of a red blood cell.Anemia means is the lack of blood and a decrease in number of red blood cells (RBCs) or less than the normal quantity of hemoglobin in the blood (Parmar et al., 2011;Ndukwu,2012). The significant reduction in the mass of circulating red blood cells and the oxygen binding capacity of the blood is diminished. Because blood volume is normally maintained at a nearly constant level, anemic patients have a decrease in the concentration of red cells hemoglobin in peripheral blood. Hemoglobin and hematocrit levels vary with the age of the individual and gender (Beutler and Waalen, 2006; Blanc et al., 1969). However, it can include decreased oxygen-binding ability of each hemoglobin molecule due to deformity or lack in numerical development as in some other types of hemoglobin deficiency. Hemoglobin is found inside RBCs normally carries the oxygen from the lungs to the capillaries. Since, all alive human cells depend on oxygen for surviving, where varying degrees of anemia can have a wide range of clinical consequences due to the hypoxia (lack of oxygen) in organs (Parmar et al., 2011).

The peripheral blood specimens were examined under the microscope to assess the size, shape, and color of the red blood cells. Normal mature red blood cell can be described as round, elastic, non-nucleated, bi-concave and has an area of central pallor, which covers about one-third of the cell. Normal mature red blood cells have an average diameter of 7.2 microns with a range of 6-9 microns (John et al., 2012). The most command anemic cases are Iron deficiency anemia, megaloblastic anemia, thalassemia, South East Asia anemia, hemolytic anemia, sickle anemia, sideroblastic, Hereditary Spherocytosis, Elliptocytosis and stomatocytosisanemia, anemia of chronic disorders (Annette, 2003). There are two general approaches that could use to identify the anemia that involved kinetic and morphologic approaches. Morphological approach is known as a peripheral smear or red blood cell morphology (Tefferi, 2003; Guyatt et al., 1992; Gjorup et al., 1986). This study focuses in morphologic approaches to extract the shape of anemic red blood cell to in the most frequent cases in Serdang hospital. In order to create a database for anemic red blood cells. There is no more literature review with comprehensive information of abnormal shapes and frequency of erythrocytes in different anemic cases. Therefore, the objective of this study was to determine the abnormal shapes and frequency rate of the erythrocytes upon the different types of anemia.

II. LITERATURE REVIEW

Shape plays a key role in image processing and pattern matching. Khot (2012) apply Artificial neural network to classify 6 type of red blood cells. Ramin(2010) classified three types of RBC in a peripheral blood smear based on morphological methods. Ruihu(2008) used depth map and surface curvature incorporated to deal with ten kinds of cell shape. Navin (2011) classify the structure of two types of red blood cells. While Thaipanich (2008) used K-means clustering (K-means) technique for robust classification of blocks in noisy images. Ramanjot(2011) apply K-mean clustering algorithm to Enhanced Image Segmentation for Liver. Rajini (2011) propose K-mean clustering algorithm to a new center initialization algorithm for measuring the initial centers of the proposed clustering algorithms. In another hand. Beham (2012) proposes an efficient K-means clustering algorithm under Morphological Image Processing. Priyadarsini(2012) applied the K-mean algorithm makes the clusters effectively forcyst area extraction from liver images. There are different types of databases. Sapana(2013) investigational nucleated red blood cells databaseover 27-mo period. Veluchamy (2012) create database includes twenty-seven features. Wheeless (1994) used 42 features for Sickle anaemia.

III. MATERIALS AND METHODS

This section includes two parts hematology and image processing; each part includes several process steps.

A.Hematology, Peripheral blood smear slides related anemic cases were obtained from hematology unite/ pathology department/ faculty of medicine/ Al- Yarmok Hospital from March to August 2018. It is through the study of 150 blood smear images for different blood disease slides includes anemic, blood cancer and Liver diseases as show in figure 1, more than 50 imagines were obtained of each slide. Morphologic abnormalities of peripheral red blood cells using microscopic examination with the oil immersion lens of well-prepare felid that red blood cells individual separated stained with Wright's stain. It has been used Olympus BX43 photo imaging microscope U-CAM D3 (Japan) in Baghdad University / Iraq

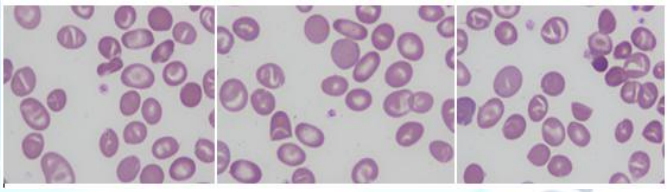

Figure 1. Different types of RBCs

B. Image processing. This part aims to extract the 100 features. There are three mine stages processes are done, data collected, pre-processing and feature extraction are illustrated in figure 2.

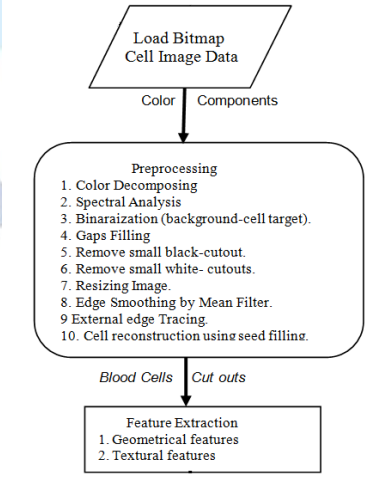

Figure 2 The graph above shows Image processing step

In the pre-processing stage several image processing operations are performed, this stage consists of ten complementary steps, starting from analyzing the color contents of the input cell image (in order to isolate background from target colors), then making image binarization to isolate target (cell) from background pixels, after that making cell segmentation using boundary tracing method.

One of the main problems we met in this work is the color variability of both the background and the cell's pixels, this variability is due to different kind of paints were used to enhance the visual appearance of the blood test samples as show in Figure 3. To handle this task, we use k-means well-known clustering algorithm to distribute the pixel's color around two dominant colors (centroids) [Jameela,2014]. One of these centroids will be very close to the dominant color of the background pixels. The implemented steps could be summarized as follows:

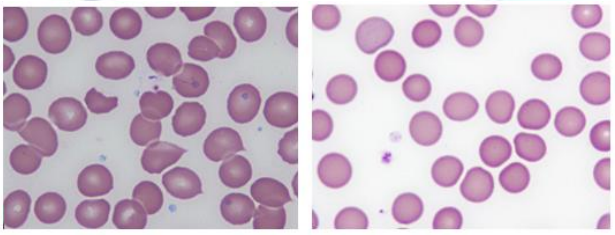

Figure3. blood smear images with color variability

Step 1: Scan the pixels lay within the strip (of width=w; say w= 5) surrounding the image boundary area, here we assume that the most of the pixels lay within this surrounding area belong to the background, so the average values of the color components of pixels belong to strip will construct the initial centroid of the background color.

Step 2: In same way the pixels in the interior region of the images are scanned and the average values of their colors are determined and considered as initial centroids for the target (cells) pixels.

Step 3: Increase the distance between the initial background and target centroids by applying the following equations:

$$R_c = R_b + \alpha(R_c - R_b) \quad (1)$$

$$G_c = G_b + \alpha(G_c - G_b) \quad (2)$$

$$B_c = B_b + \alpha(B_c - B_b) \quad (3)$$

where,

(Rb,Gb,Bb) are the red, green, and blue components of the background centered.

α: is the separation parameter

Step 4: Distribute all the image pixels among the two clusters (background or cell) according to their distances from the two corresponding centroids (background or cell). The following distance criteria (i.e., nearest distance) was utilized:

```
if Distance (Pixel color, Background centroid)
   < Distance (Pixel color, Cell centroid) then
        Set Pixel Index = Background
else
        Set Pixel Index = cell
```

where,

$$Distance(PixelColor, BackGroundCentroid) = \sqrt{(R-R_b)^2+(G-G_b)^2+(B-B_b)^2} \quad (4)$$

$$Distance(PixelColor, CellsCentroid) = \sqrt{(R-R_c)^2+(G-G_c)^2+(B-B_c)^2} \quad (5)$$

(R,G,B) are the color components of the tested pixel.

Step 5. Determine the average color components for all pixels indexed as background using the following equations:

$$R'_b = \frac{1}{n_b}\sum_{i\in b} R_i \;;\; G'_b = \frac{1}{n_b}\sum_{i\in b} G_i \;;\; B'_b = \frac{1}{n_b}\sum_{i\in b} B_i \quad (6)$$

where

b: is the set of image pixels indexed as background.

nb: is the size of the set (b).

Step 6. Determine the average color components for all pixels indexed as target (or cell) using the following equations:

$$R'_c = \frac{1}{n_c}\sum_{i\in c} R_i \;;\; G'_c = \frac{1}{n_c}\sum_{i\in c} G_i \;;\; B'_c = \frac{1}{n_c}\sum_{i\in c} B_i \quad (7)$$

where

c: is the set of pixels indexed as cell pixel.

nc: is the length (or size) of the set (c).

Step 7. Compare the difference (D) between the vectors (R'_b, G'_b, B'_b) and (Rb,Gb,Bb), and between the (R'_c, G'_c, B'_c) vectors and (Rc,Gc,Bc) by using the following distance measure:

$$D^2 = \frac{1}{6}[(R'_b - R_b)^2 + (G'_b - G_b)^2 + (B'_b - B_b)^2 + (R'_b - R_b)^2 + (G'_b - G_b)^2 + (B'_b - B_b)^2] \quad (8)$$

$$R_b = R'_b \;;\; G_b = G'_b \;;\; B_b = B'_b \quad (9)$$

$$R_c = R'_c \;;\; G_c = G'_c \;;\; B_c = B'_c$$

Step 9.If D is greater than a pre-defined threshold value (Dmin) then

repeat steps (4 to 8), otherwise exit.

This processes are helpfully to get the best segmentation process. The third steps in image processing is feature extraction: In this stage a set of discriminating 60 geometrical and (40) textural and color features, a large set of geometrical cells using textural features have been used (such as, Packing, Fourier descriptor, Absolute moments,

and color moments) these features are determined from the input color image using only the pixels labeled in the constructed binary image as (white) pixels. The features (Jameela, 2013). The determined features are grouped all together to produce the feature vector.

- Packing;

$$Packing = \frac{(No\ of\ External\ Boundary\ Pixels)^2}{Total\ No\ of\ Cell\ Pixels}$$

- Fourier descriptor;

$$F(n) = \frac{T}{2\pi^2 n^2}\sum_{i=1}^{M}\frac{\Delta x(i)}{\Delta t(i)}\left[\cos\left(\frac{2\pi}{T}\sum_{j=1}^{i}\Delta t(j)\right) - \cos\left(\frac{2\pi}{T}\sum_{j=1}^{i-1}\Delta t(j)\right)\right] \quad (10)$$

n=1...4

Where,

$\{x_0, x_1, x_2, ..., x_M\}$ are the $x-$coordinate of external boundary points.

$\{x_0, x_1, x_2, ..., x_M\}$ are the $y-$coordinate of external boundary points.

M is the number of boundary points.

$\Delta x(i) = x(i) - x(i-1)$

$\Delta y(i) = y(i) - y(i-1)$

$\Delta t(i) = \sqrt{(\Delta x(i))^2 + (\Delta y(i))^2}$

$T = \sum_{i=1}^{M}\Delta t(i)$

- Absolute Moment

$$aMom(n) = \frac{1}{k}\left|\sum_{i\in c}\left|(x_i' + jy')^n\right|\right| \quad (11)$$

n =1,2,3,4

Where,

$$x_i' = \frac{1}{L}\left[(x_i - \bar{x})\cos\theta - (y_i - \bar{y})\sin\theta\right]$$

$$y_i' = \frac{1}{L}\left[(x_i - \bar{x})\sin\theta + (y_i - \bar{y})\cos\theta\right]$$

IV. BLOOD CELL IMAGE DATASET

Abnormal red blood cell dataset includes 100 features for 30 classes. These features are 60 geometric and 40 texture and color. These features arrangement in Excel file formatting. Figure 4 show part of Excel file that includes features.The features in Excel file was transfer it to Comma-separated values (CSV) file. After that, created RBFF file format to use in Weka softwear before start modeling classification inorder to demonstrate the accuracy of the features.

Figure 4. the part of Excel features file

f0	f1	f2	f3	f4	f5	f6	f7	f8
29250	10.41723	0.202696	-0.07882	0.4815	-0.4899	-0.0482	-1E-05	-0.00474
24029	13.7116	0.153264	-0.3838	0.2517	-0.3948	-0.2395	-0.04998	-0.03006
33843	10.63735	0.258044	-0.04239	0.5027	-0.5025	-0.0356	0.001098	-0.00255
33714	10.53616	0.433309	-0.04151	0.4989	-0.5029	-0.05477	-0.00269	-0.00465
30887	10.59294	8.06E-02	-0.1093	0.4826	-0.4982	-0.05659	-0.00199	-0.0016
30651	10.63721	2.75E-02	-0.08957	0.4935	-0.4946	-0.07871	-0.00108	-0.00118
31578	10.3974	0.548984	-0.0264	0.5001	-0.4995	-0.00297	-0.00072	-0.00042
31534	10.52122	0.841784	-0.03661	0.4987	-0.4996	-0.04386	-6.5E-05	-0.00272
30490	10.46983	7.39E-02	-0.02355	0.4991	-0.4982	0.005681	0.002534	0.001333
31578	10.3974	0.548984	-0.0264	0.5001	-0.4995	-0.00297	-0.00072	-0.00042
32146	11.31117	0.470259	-0.1673	0.4604	-0.4874	-0.1003	-0.00113	-0.00172
27234	12.43754	0.362485	-0.287	0.3558	-0.4574	-0.1256	-0.03769	0.01808
33278	11.29181	0.974458	0.01398	0.4939	-0.4935	-0.03833	0.0139	-0.00974
35109	10.98411	0.558119	0.01787	0.4907	-0.4927	-0.08876	-0.00073	-0.00115
29517	11.75326	0.546963	-0.1438	0.4408	-0.446	-0.1451	-0.00369	0.006064
36763	13.13889	0.670416	0.03884	0.4754	-0.4671	-0.05358	-0.00255	-0.01344
32306	10.77509	0.494637	-0.07794	0.4843	-0.4892	-0.05901	-0.0022	-0.00226
31750	13.10315	0.444176	0.2186	0.4299	-0.4617	0.1012	-0.01338	0.009141
30382	10.39576	0.112156	-0.02571	0.4985	-0.4987	-0.01357	-0.00249	0.003974
32450	13.46444	0.524779	0.1198	0.4634	-0.472	0.07918	-0.00548	-0.00336

V. RESULT

These features are examined with different classification machine learning algorithms (ML), such as SVM, ANN, Nearest Neighbor(NN) and Zero R. Table 1show the classification accuracy for the algorithms.

Table 1. classification ML algorithms

Algorithms	Correctly C.	Incorrectly C.	Accuracy
SVM	944	56	94.4%
ANN	910	90	91%
NN	900	100	90%
ZeroR	890	110	89%

The accuracy for each algorithm are devaluated to improve the accurate features using precision recall and Root mean squared error based on Confusion Matrix, as show in Table 2 bellow. The Sensitivity is the true positive rate and Specificity is the true negative rate.

Table 2. the evaluation of classification algorithms

	Algorithms			
	SVM	ANN	NN	Zero.R
Sensitivity	94.4	91	90	89
Specificity	95.1	90.6	90	89.9
MRSE	0.236	0.269	0.298	0.387

VI. CONCLUSION

In this paper, pre-processing stage several image processing operations are performed. The main problems we met in this work is the color variability of both the background and the cell's pixels. This problem solved with k-means clustering algorithm to distribute the pixel's color around two dominant colors (centroids).100 geometrical, color and texture features obtained for 1000 individual RBCs images. Different machine learning classification algorithms are

used to test these features; the algorithms achieve high accuracy more than 94%. The algorithms evaluated and achieved more than 94% Sensitivity, 95% Specificity and leass than 0.38 Root mean squared error. These result conclude that the algorithms have high ability of distinguish between the true negative and true positive, the fact that the features are very accurate in describing RBCs

REFERENCES

[1] Deyi Xu. 2012. "Study of damage to red blood cells exposed to different doses of γ-ray irradiation" Blood Transfus. July; 10(3): 321–330.

[2] Blanc B, Finch CA,1968.Hallberg L, et al. Nutritional anaemias. Report of a WHO Scientific Group. WHO Tech Rep Ser.;405: 1-40.

[3] Beutler E and Waalen J.2006. The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration. Blood March 1, vol. 107 no. 5 1747-1750.

[4] Ndukwu, G. U., &Dienye, P. O. (2012). Prevalence and socio-demographic factors associated with anemia in pregnancy in a primary health center in Rivers State, Nigeria. African Journal of Primary Health Care & Family Medicine, 4(1).

[5] Parmar K, Mithilesh Patel, Parthiv Chauhan.2011. Review: A Review On Anaemia International Journal of Comprehensive Pharmacy, 02; 11: 1-6.

[6] Tefferi A. 2003. Anemia in Adults: A Contemporary Approach to Diagnosis. Mayo Clin Proc. 78:1274-1280.

[7] Gjorup T, Bugge PM, Hendriksen C, Jensen AM. 1986. A critical evaluation of the clinical diagnosis of anemia. Am J Epidemiol.;124:657-665.

[8] Guyatt GH, Oxman AD, Ali M, Willan A, McIlroy W, Patterson.1992. Laboratory diagnosis of iron-deficiency anemia: an overview correction appears in J Gen Intern Med.;7:423]. J Gen Intern Med. 1992;7:145-153.

[9] Annette Carley. 2003.Anemia: When Is it Not Iron Deficiency?:PediatrNurs 29(3):205-211,. © Jannetti Publications, Inc.

[10] John M. Higgins; Ramachandra R. Dasari; Subra Suresh;YongKeun Park. 2012. Anisotropic light scattering of individual sickle red blood cells J. Biomed. Opt. 17(4), 040501 (Apr 05).

[11] Giovanna Tomaiuolo.2011." Start-up shape dynamics of red blood cells in microcapillary flow" 6. http://dx.doi.org/10.1016/j.mvr.

[12] Ciccoli L. C. De Felice, E. Paccagnini, S. Leoncini, A. Pecorelli, C. Signorini, G. Belmonte, G. Valacchi, M. Rossi and J. Hayek. 2011 "Morphological changes and oxidative damage in Rett Syndrome erythrocytes".

[13] Warhurst DC, Williams JE .1996. "Laboratory diagnosis of malaria". J ClinPathol 49 (7): 533–38. doi:10.1136/jcp.49.7.533. PMC 500564. PMID 8813948,(1996).

[14] ElianeGluckman, Hal E. Broxmeyer, Arleen D. Auerbach, Henry S. Friedman.2010 Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. Cellular Therapy and Transplantation (CTT), Vol. 2, No. 7 ;2:e.000079.01.doi:10.3205/ctt-e000079.01.

[15] Carl R. Kjeldsberg. 1995. Practical diagnosis of hematologic disorders. Chapter one diagnostic anemia Sherrie Perkins, ASCP Press, - Medical – 3-14 pages.

[16] M. Lindenbaum and M. Fischer and A. M. Bruckstein.1994. "On Gabor's contribution to image-enhancement," Pattern Recognition, vol. 27, pp.1-8.

[17] TanapholThaipanich and C. -C. Jay Kuo. 2010. An Adaptive Nonlocal Means Scheme for Medical Image Denoising Proc. SPIE 7623, Medical Imaging 2010: Image Processing, 76230M (March 12); doi:10.1117/12.844064.

[18] Farahanirad, H., et al. 2011. "A Hybrid edge detection algorithm for salt-and-pepper noise." proceedings of the IMECS.

[19] Soltanzadeh, Ramin, and Hossein Rabbani. 2010. "Classification of three types of red blood cells in peripheral blood smear based on morphology." Signal Processing (ICSP), IEEE 10th International Conference on. IEEE.

[20] Wang, Ruihu, and Brendan McCane. 2008. "Red Blood Cell Classification through Depth Map and Surface Feature." Computer Science and Computational Technology. ISCSCT'08. International Symposium on. Vol. 2. IEEE, 2008.

[21] Jambhekar, Navin D. 2011."Red Blood Cells Classification using Image Processing." ISSN: 2249-7846 Science Research Reporter 1(3): 151-154, Nov.

[22] Thaipanich, Tanaphol, Jay Kuo.2010. "An adaptive nonlocal means scheme for medical image denoising." SPIE Medical Imaging. International Society for Optics and Photonics.

[23] Rajini, N. Hema, and R. Bhavani. 2011 "Enhancing k-means and kernelized fuzzy c-means clustering with cluster center initialization in segmenting MRI brain images." Electronics Computer Technology (ICECT), 2011 3rd International Conference on. Vol. 2. IEEE.

[24] Beham, M. P., and A. B. Gurulakshmi. 2012."Morphological image processing approach on the detection of tumor and cancer cells." Devices, Circuits and Systems (ICDCS), International Conference on. IEEE.

[25] Priyadarsini, S., and D. Selvathi. 2012."Survey on segmentation of liver from CT images." Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on. IEEE.

[25] Zhang, Wei, et al. 1998. "Shape-based indexing in a medical image database."Biomedical Image Analysis, 1998. Proceedings. Workshop on. IEEE.

[26] Kim, KyungSu, et al. 2000. "Automatic classification of cells using morphological shape in peripheral blood images." SPIE proceedings series. Society of Photo-Optical Instrumentation Engineers.

[27] S. T. khot ,dr.prasadr.k. " image analysis system for detection of red blood cell disorders using artificial neural network" (ijert) vol. 1 issue 5, july - 2012 issn: 2278-0181.

[28] Zajicek, G., and M. Shohat.1983. "On the classification of nucleated red blood cells."Computers and biomedical research 16.6 553-562.

[29] Ali, J., Ahmad, A. R., George, L. E., Der, C. S., & Aziz, S. 2013. Red blood cell recognition using geometrical features. International Journal of Computer Science Issues (IJCSI), 10(1), 90.

[30] Connelly, T. J.2014. Mapping Aspect Ratios in the Age of High-Definition Television.Popular Communication, 12(3), 178-193.

[31] Jameela,A,Alkrimi., George, L. E., Suliman, A., Ahmad, A. R., & Al-Jashamy, K. 2014. Isolation and Classification of Red Blood Cells in Anemic Microscopic Images. World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering, 8, 727-730.