International Journal Of Advanced Communication Technology (IJACT)
Volume 4, 2024 [1-18]

Efficient Hexagonal Gridding Method for DNA Microarray Image

Maziidah Mukhtar Ahmad [1] and Asral Bahari Jambek [2]

[1] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Arau, 02600, Perlis, Malaysia
[2]Centre of Excellence for Micro System Technology (MiCTEC), Universiti Malaysia Perlis, Arau, 02600, Perlis, Malaysia

Corresponding author: Maziidah Mukhtar (e-mail: jidan8189@gmail.com).

ABSTRACT

In genetics, a deoxyribonucleic acid (DNA) microarray is a useful instrument that is frequently used to track thousands of genes' expression levels simultaneously. For DNA microarray, gene expression is done through microarray spot gridding, segmentation and intensity extraction. The gridding processes identify each microarray spot location and supply their coordinates for the spot. Many microarray technologies arrange their microarray spots in in rectangular fashion. However, a hexagonal grid arrangement is also being used to increase the density of the spot per unit area. While gridding microarray arranged in rectangular fashion is straightforward, gridding microarray arranged in hexagonal fashion posed its own challenge due to the irregular structure of the spot location. To solve this problem, this paper discusses two proposed methods to perform gridding for microarrays arranged in a hexagonal fashion, namely alternate-column and diamond-shape methods. These methods have been evaluated in nine different benchmark images that represent best, typical and worst-case images. Based on our experiments, the proposed diamond-shape and alternate-column methods achieved an average accuracy of 93.6% and 92.1%, respectively. In terms of computational load, the diamond shape consumes three times less computational loads as compared to the alternate-column.

KEYWORDS *DNA microarray, gridding, hexagonal, accuracy, computational time*

I. INTRODUCTION

Many people believe that microarrays will bring about the next great advancement in molecular biology. It makes it possible for scientists to do genomic-scale analyses of genes, proteins, and other biological molecules. When analyzing the levels of genomic expression in physiologic and pathologic phenomena, microarrays are incredibly effective tools. A common tool in many genomic research labs these days is the microarray. Its revolutionary approach to biological research is the reason for its popularity. Tens of thousands of genes can now be studied at once by scientists, rather than working on a gene-by-gene basis. Due in large part to the future potential of this technology as well as the sheer technical promise of this methodology, many early microarray experiments have been published in high-impact journals.

No other technology in the history of biology has employed as much technological sophistication, brought together knowledge from various disciplines, or produced such a detailed view of the cell as microarray analysis [1] & [2]. A microarray is made up of several tiny components, or spots, arranged in columns and rows. In the early 1990s, Schena and colleagues at Stanford University's Davis Laboratory and Stanford Biochemical Department developed microarrays. A crucial component of microarray research is image analysis, which may have a significant influence on identifying genes with differential expression. Three tasks are involved in image processing for microarray images: spot gridding, segmentation and intensity extraction. Gridding techniques based on the distribution of pixel intensities are crucial for the

*jidan8189@gmail.com

analysis of microarray experiments. By extracting the foreground and background intensities of red, green, and yellow for each spot, image analysis on microarray images aims to ascertain the degree of hybridization of each spot. Microarrays can be used to identify genes involved in a specific disease by comparing gene expression in normal and abnormal cells.

One of the most crucial phases in the analysis of microarray images is gridding. Using the different segmentation rules, the located spot region can be separated into the background region and the foreground spot region. In actuality, there is background intensity surrounding the location. Before extracting image intensities, an appropriate global or local background intensity correction technique should be applied. Following spot segmentation, flagging is the process to be carried out to eliminate or label low-quality intensity features. Eventually, the clustering of genes with differential expressions will be made possible by the intensity values that were generated. Preprocessing reduces common artefacts in images, improving the accuracy of gridding, segmentation, and intensity extraction. Thus, this phase cannot be avoided. Therefore, we concentrated on using an automated preprocessing technique along with gridding the images. By automating this step of the process, high throughput is made possible [3].

In a microarray experiment, the hybridized array images are captured to measure the red (Cy5) and green fluorescence (Cy3) intensities for each spot on the glass slide. These fluorescent intensities correspond to the level of hybridization of the two samples of the DNA sequences spotted on the slide [4]. The fluorescent intensities are stored as 16-bit images. Figure 1 shows an ideal microarray image [5]. Some of the important characteristics that constitute an ideal microarray image include the number of rows and columns in each sub-array being the same, the spots are not rotated or inclined in a horizontal or vertical direction, the spacing between sub-arrays and between spots is regular, all of the spots have the same sizes and shapes, the microarray slide is free of dust and other contaminants, the background intensity should remain constant throughout the image and the spots should only reflect the true measures of fluorescence intensities of a corresponding dye of interest [6].

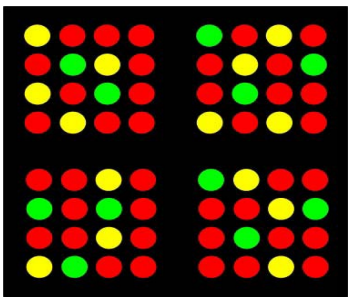


Figure 1: Illustration of an ideal microarray image [5]

II. VARIATION OF MICROARRAY SPOT ARRANGEMENT

The probes are arranged on the slide in a rectangular grid by the majority of microarray manufacturers. In a rectangular grid, each probe is positioned at precise, fixed intervals along both the X and Y axes, creating a matrix of spots, as shown in Figure 2. The first step in processing an image is called gridding, and it involves finding the coordinates of each spot. The first step in this process is image reading; for this work, an image in JPEG format was used. The image's selected

region of interest is then cropped [7]. The cropped image is then divided into three image planes: red, green, and yellow. In addition, the image is transformed into a grayscale version which will be utilized for gridding. The ideal scenario would be for every spot to be consistently and periodically located. In true terms, however, the spots vary in size and intensity. After that, the binary image is processed using the Projection Profile Method. To decrease noise and measure spot distance, both horizontal and vertical profiles have been used [8]. The profile's negative peaks are found to line up with the vertical grid lines' locations. There are noise and imperfections in the original image, which could lead to redundant or absent grid lines. Consequently, de-noising and projection profile refinement are needed in order to correctly grid the image. Top hat filtering was used to enhance the mean horizontal profile derived from the autocorrelation profile's peak values [9]. It enhances the details that would otherwise be hidden in low contrast regions. Segmenting peaks or using these individuals as a vertical and horizontal separator comes next, after all background noise has been eliminated [10]. It is possible to determine the distance between two spots using the mean profile projection value. The gridline or bonding box can be drawn correctly by locating the midpoint between the two spots on the left and right, as well as the top and bottom. The horizontal centres of the spots line up with the centroids of the peaks. These are the feature extraction tasks that can be carried out using region props [11]. Bounding boxes can be used as regions to create a rectangular grid by addressing each spot independently. For ease of use, each bounding box's dimensions, coordinates, and position were tabulated into 4-column matrix regions of interest, or ROIs.

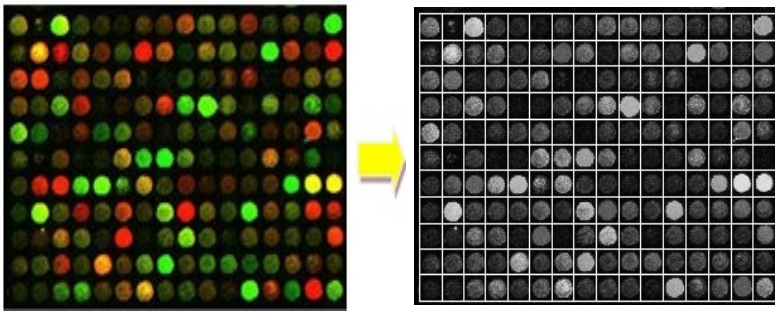


Figure 2: Rectangular spot arrangement

Other than rectangular gridding, hexagonal gridding is also being used to arrange microarray spots. Authors in [10] state that Illumina developed the Illumina Bead Array Technology using a hexagonal structure in response to the need to increase the number of probes in a single microarray experiment. According to [10] the hexagonal grid has an advantage over the rectangular grid for a microarray slide's structure because it permits the packing of more probes in one location.

Figure 3 illustrates the proposed method of gridding for microarray arranged in hexagonal arrangement. Figure 3. (a) shows a portion of the raw image, and Figure 3. (b), (c), and (d) show the steps suggested in [12], which are the hybridized spot detection, the non-hybridized spot detection, and gridding, respectively. In this method, first, the suggested technique uses image projections to divide the blocks. The projections' 1D signal undergoes several processing stages to emphasize the valleys in it. Then, the image is divided into discrete blocks using the coordinates of these valleys. The method uses the Growing Concentric Hexagon (GCH) algorithm to detect

every spot in the image, including the dark spots. The algorithm simultaneously verifies the spots that have already been detected and estimates the positions of the non-hybridized spots. Using the locations of its neighbours, the algorithm finds multiple spots on the contour of a growing hexagon with each iteration. Furthermore, the locations of the detected spots are optimized using the intensities of the previously identified high-intensity spots in each algorithm iteration. Lastly, the area assignment of each spot in the image is performed using the Voronoi diagram.

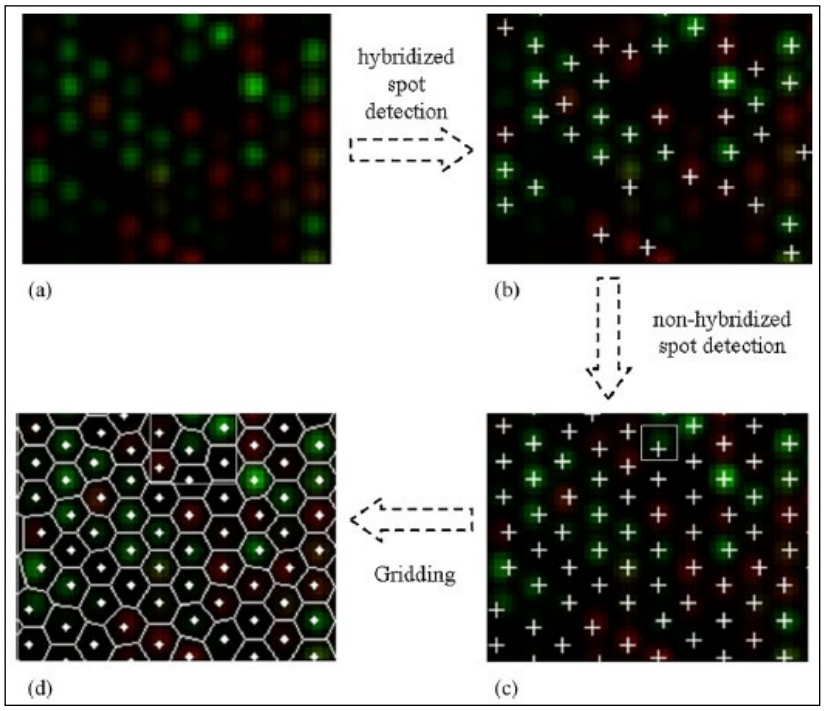


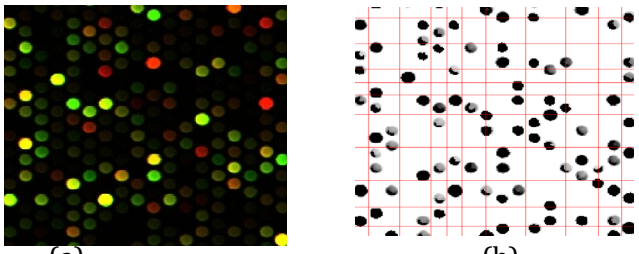
Figure 3: Findings from the suggested approach: (a) the raw image, (b) the centres (white crosses) of the high-intensity spot, which are detected by the hybridized spot detection step, (c) all the non-hybridized spots which are detected by the GCH algorithm, and (d) the application of the Voronoi diagram for the detected centres. [11]

III. METHODOLOGY

A Proposed Gridding Method for Hexagonal Spot Arrangement Using Alternate-column

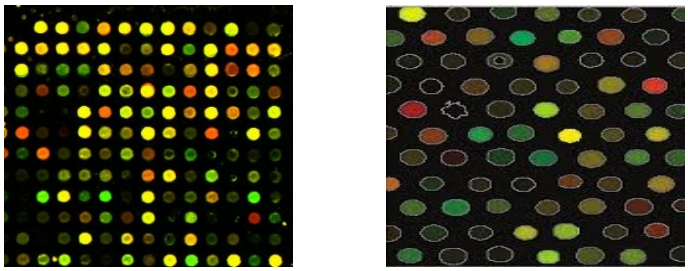
Figure 4 shows the result of applying a rectangular grid to a hexagonal-arranged microarray spot image. From the figure, it is obvious that the rectangular gridding cannot be used for this hexagonal spot arrangement because each row and column in the hexagonal arrangement has a non-parallel spot position. Finding the horizontal profile, as discussed in the previous section, while calculating the difference in distance between two spots is a challenging task for hexagonal spot arrangement. In this section, a new gridding algorithm for microarray spot

arrangement in hexagonal arrangement will be proposed. Figure 5 shows a comparison between microarray spots arranged in a square grid and microarray spots arranged in a hexagonal grid. Hexagonal gridding is not widely discussed and very few papers discuss hexagonal gridding for microarrays such as in [13]. Furthermore, the current proposed method for hexagonal gridding requires a high computational load because the process to get the grid of each spot is more complex which the spot will grid one by one. For sure this process requires a high computational load [14]. In order to reduce the computational load to grid microarray spot arranged in a hexagonal fashion, two gridding algorithms are proposed in this work, namely the alternate-column method and the diamond-shape method.



(a) (b)

Figure 4: DNA microarray image (a) hexagonal gridding from Genepix
(b) Ractanguar gridding



(a) (b)

Figure 5: DNA microarray image (a) square gridding from ScanAlyze and
(b) hexagonal gridding from Genepix

Figure 6 shows the flow chart of the proposed alternate-column method. The step-by-step image and the related task performed in each step are shown in Figure 7. In this proposed method, the estimating spot spacing is modified to take into account the shape of the spot arrangement. During estimating the spot spacing, the hexagonal microarray spots (Figure 7(a)) are first profiled horizontally (Figure 7 (b)). From this profile, the vertical grid can be created for the microarray images as shown in Figure 7 (c). Based on the vertical gridding, the microarray spot is grouped column by column and each column is numbered. Once labelled, the microarray image will be re-composed where all odd-numbered columns (Figure 7 (d)) will be grouped to form a new sub-image, and all even columns numbered (Figure 7 (g)) will be grouped to form another sub-image. This results in two microarray sub-images that are arranged equivalent to the square grid.

Based on the recomposed images, each of the sub-images can be treated as a normal square grid. Next, the vertical gridding can be performed on each recomposed image (Figure 7 (f) and (i)). Once the vertical gridding has been obtained for each sub-image, we can rearrange the odd and even strip back to the original position while maintaining the vertical grid information obtained previously (Figure 7 (j)).

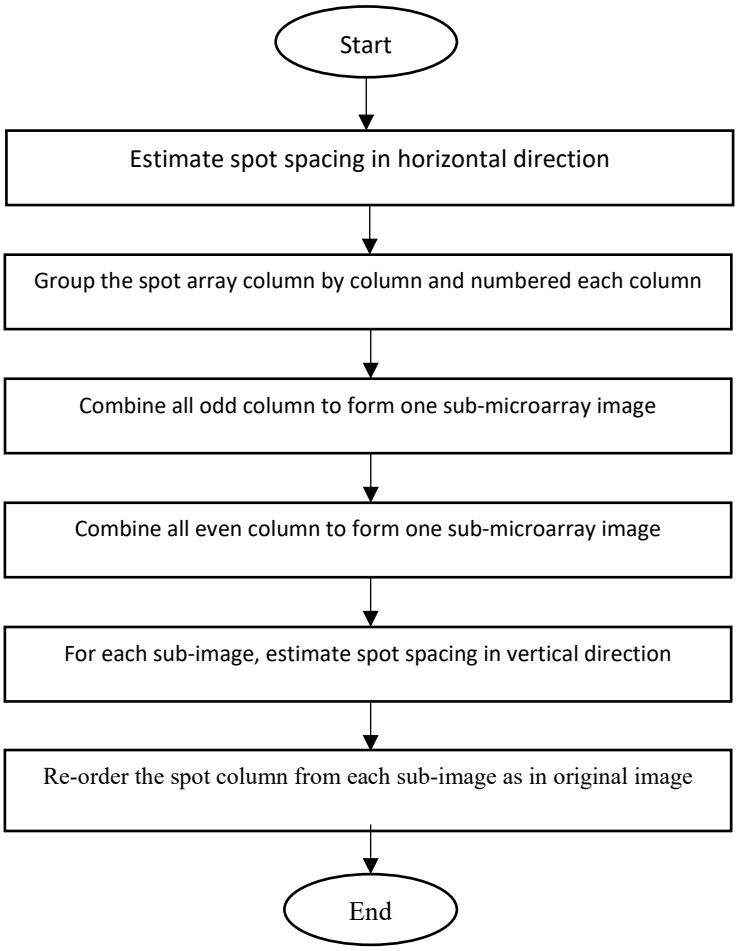
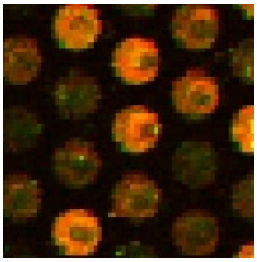
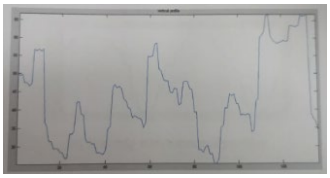


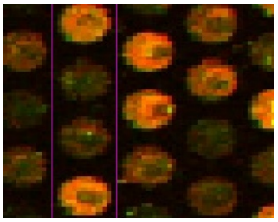
Figure 6: Modified estimate spot spacing for alternate-column method



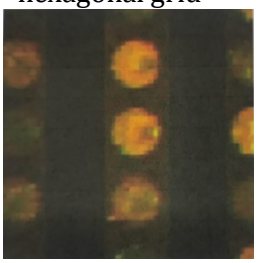
(a) Original hexagonal grid



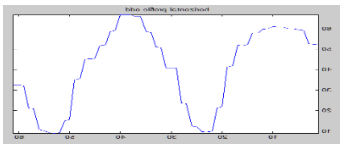
(b) horizontal profile of the original image



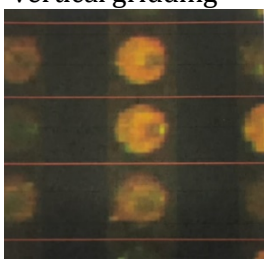
(c) original image with vertical gridding



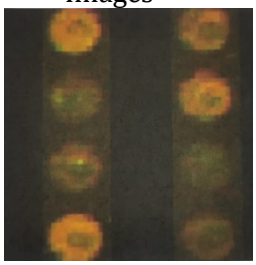
(d) Odd strip Images



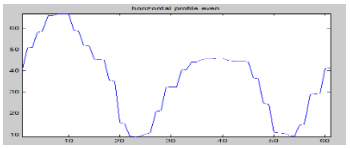
(e) Odd strip with vertical profiling



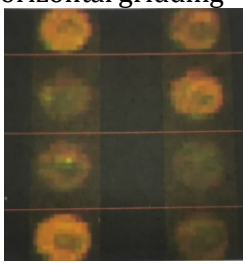
(f) Odd strip with horizontal gridding



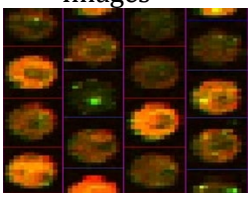
(g) even strip Images



(h) even strip with vertical profiling



(i) Even strip with horizontal gridding



(j) Recombine image

Figure 7: Alternate-column method gridding process (a) Original hexagonal grid (b) horizontal profile of the original image (c) original image with vertical gridding (d) Odd strip Images (e) Odd strip with vertical profiling (f) Odd strip with horizontal gridding (g) even strip Images (h) even strip with vertical profiling (i) Even strip with horizontal (j) Recombine image

Figure 8 shows the flow chart of the second proposed method, called diamond-shape gridding. Figure 9 shows an example of diamond-shaped gridding. In this method, the horizontal estimate spot spacing is calculated first prior to creating the vertical gridding (Figure 9 (a)). Each vertical gridding will be numbered according to their position from left to right. Next, a diagonal line is created to match the hexagonal spot arrangement as shown in Figure 9(b). A predetermined diagonal line of 30^0 is taken to match the hexagonal arrangement pattern. The final gridding is

obtained by overlapping the vertical gridding and the diagonal gridding. The intersection between these lines will define the location of each spot.

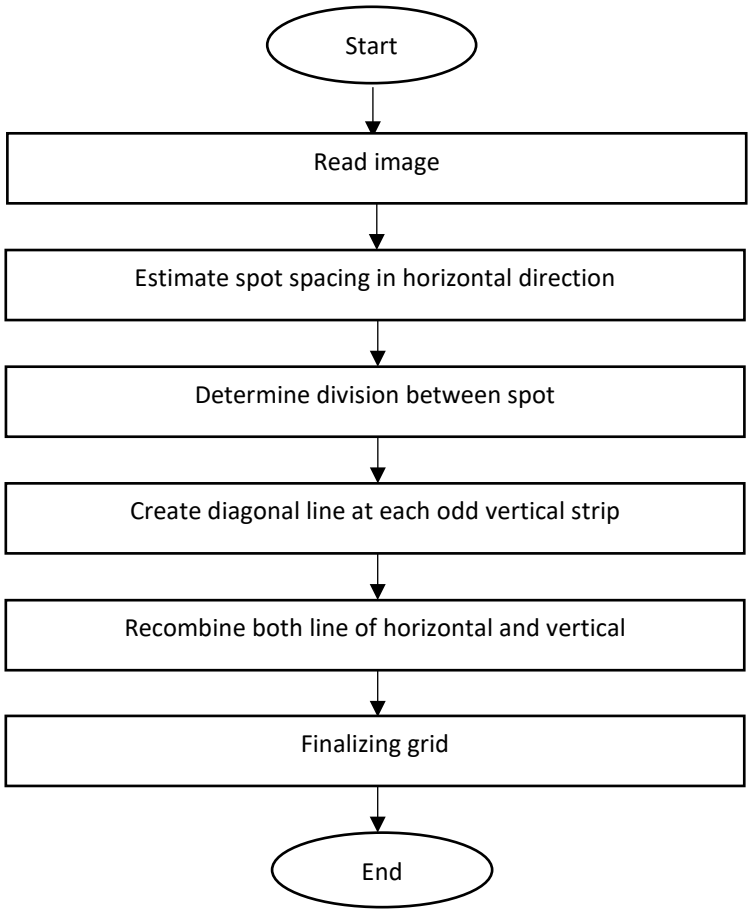
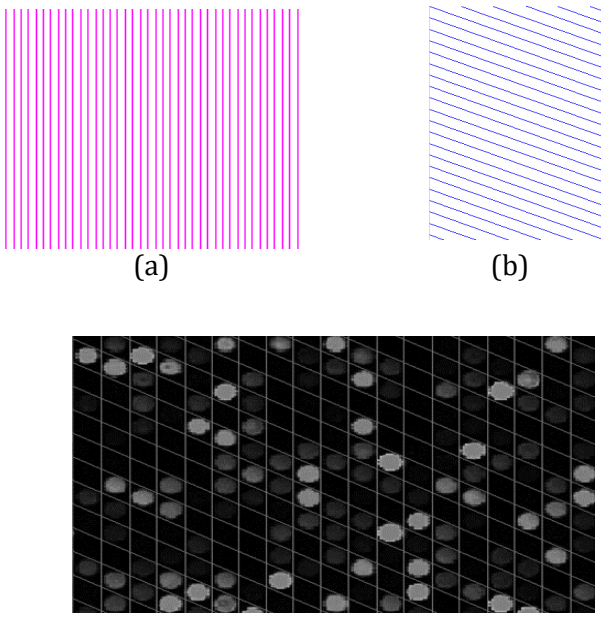


Figure 8: Flow chart of diamond gridding.



(a) (b)

(c)

Figure 9: Diamond-shape gridding process (a) vertical grid (b) diagonal grid (c) final diamond-shape gridding

IV. RESULT AND DISCUSIION

The proposed algorithms are evaluated on nine different benchmark DNA microarray images arranged in a hexagonal fashion. Images P, Q and R represent best-case images, while images S, T and U represent typical case images. The worst-case image is represented by images V, W and X. These nine images are shown in Figure 10. In this work best case is the spots are located in their geometry coordinates. In the typical case images, microarray spots are not really clear and has some missing spots due to inconsistent background intensities. The worst case image represents microarray images that contains many missing spots and some spillage of liquids or acids during the process of preparing the microarray. The used microarray images are downloaded from the Stanford Microarray Database (SMD)[15].

Table 1 (a) and (b) show the accuracy results of performing hexagonal gridding using alternate-column gridding and diamond-shape gridding. From the table, the proposed methods give almost similar average accuracy of 93.6% and 92.1% for diamond-shape and alternate-column, respectively. For different image categories, both methods have almost similar gridding accuracy, with diamond-shaped gridding slightly outperforming the alternate-column approach by 2%.

Figure 11 and Figure 12 show the gridding result for diamond-shape and alternate-column methods for all benchmark images. From the figures, the alternate-column method cannot perform horizontal griding accurately if there are any weak or missing microarray spots. This is because the horizontal gridding depends on the availability of the spots in each even and odd column of the microarray to calculate the horizontal profile. Any missing spot will affect the horizontal profile calculation. On the other hand, diamond-shaped gridding can perform better in this situation since the diagonal lines are predetermined according to the structure of the microarray. Any missing spot will not affect the diagonal line in the diamond-shape method.

Figure 13 and Figure 14 show the gridding result for the proposed algorithm using a diamond shape and alternate column in more detail. Gridding alternate columns perform the same initial vertical gridding method. For the alternate-column method, the horizontal gridding is performed by profiling the generated sub-image. Thus, any noise in the image will affect the horizontal gridding. For diamond-shaped gridding, the diagonal line is created by taking 30^0 angles from the vertical gridding. The diagonal line is predetermined based on the hexagonal shape of the spots. Thus, removing any missed alignment due to noise.

Generally, images of the worst-case scenario exhibit lower gridding accuracy compared to other image types. This reduction in accuracy stems from the characteristics of spots with insufficient intensity, as illustrated in Image X. The low intensity of these spots complicates the distinction between the spot itself and the background. This is crucial for calculating the horizontal profile value and determining spot spacing which is important for the gridding process. Moreover, the variation in spot sizes further diminishes accuracy.

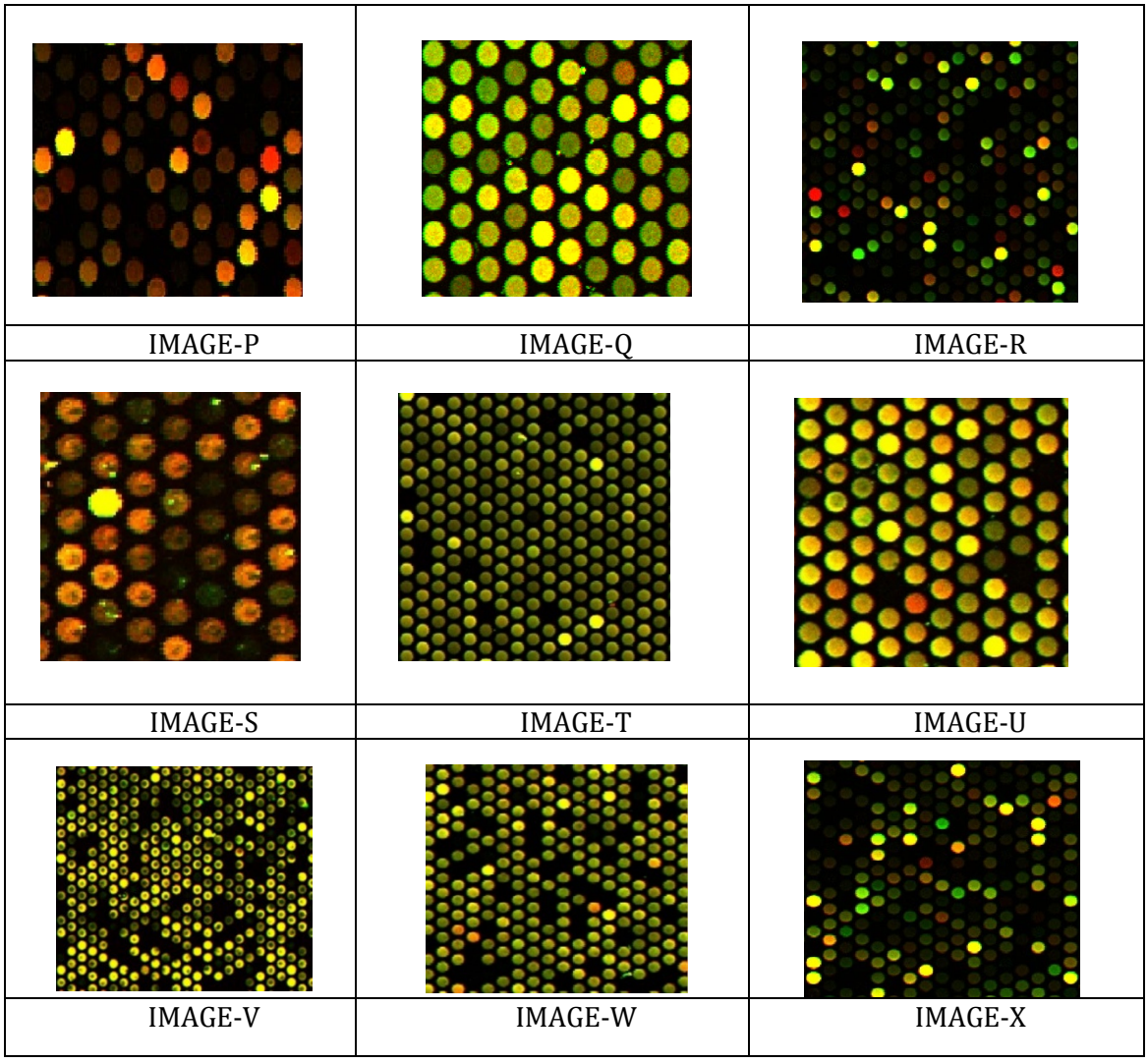


Figure 10: Nine benchmark images represent the best, typical and worst-case images.

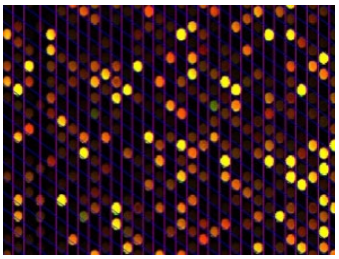
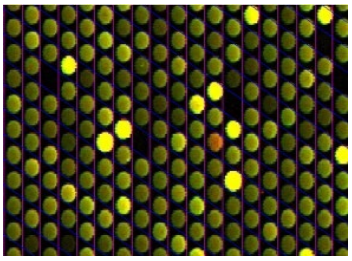
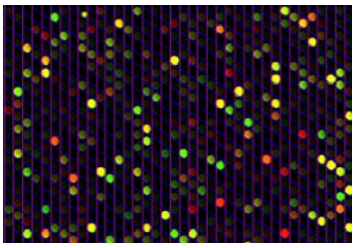
		

IMAGE-P	IMAGE-Q	IMAGE-R

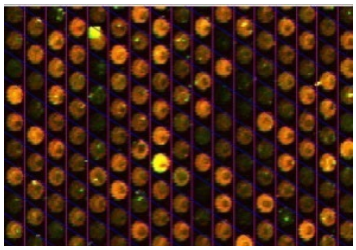
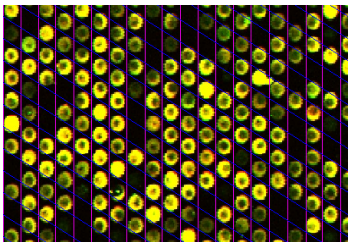
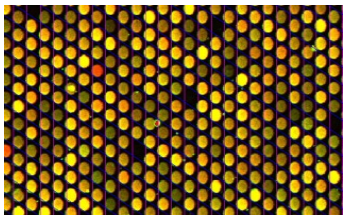

		

IMAGE-S	IMAGE-T	IMAGE-U

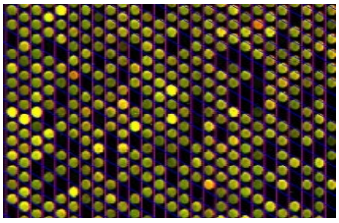
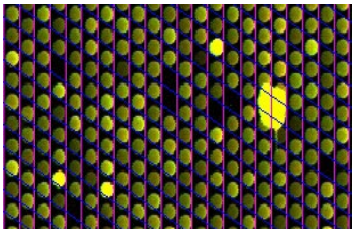
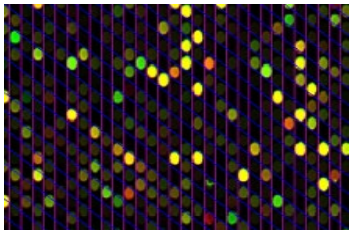

		

IMAGE-V	IMAGE-W	IMAGE-X

Figure 11: Results of hexagonal gridding using diamond-shape

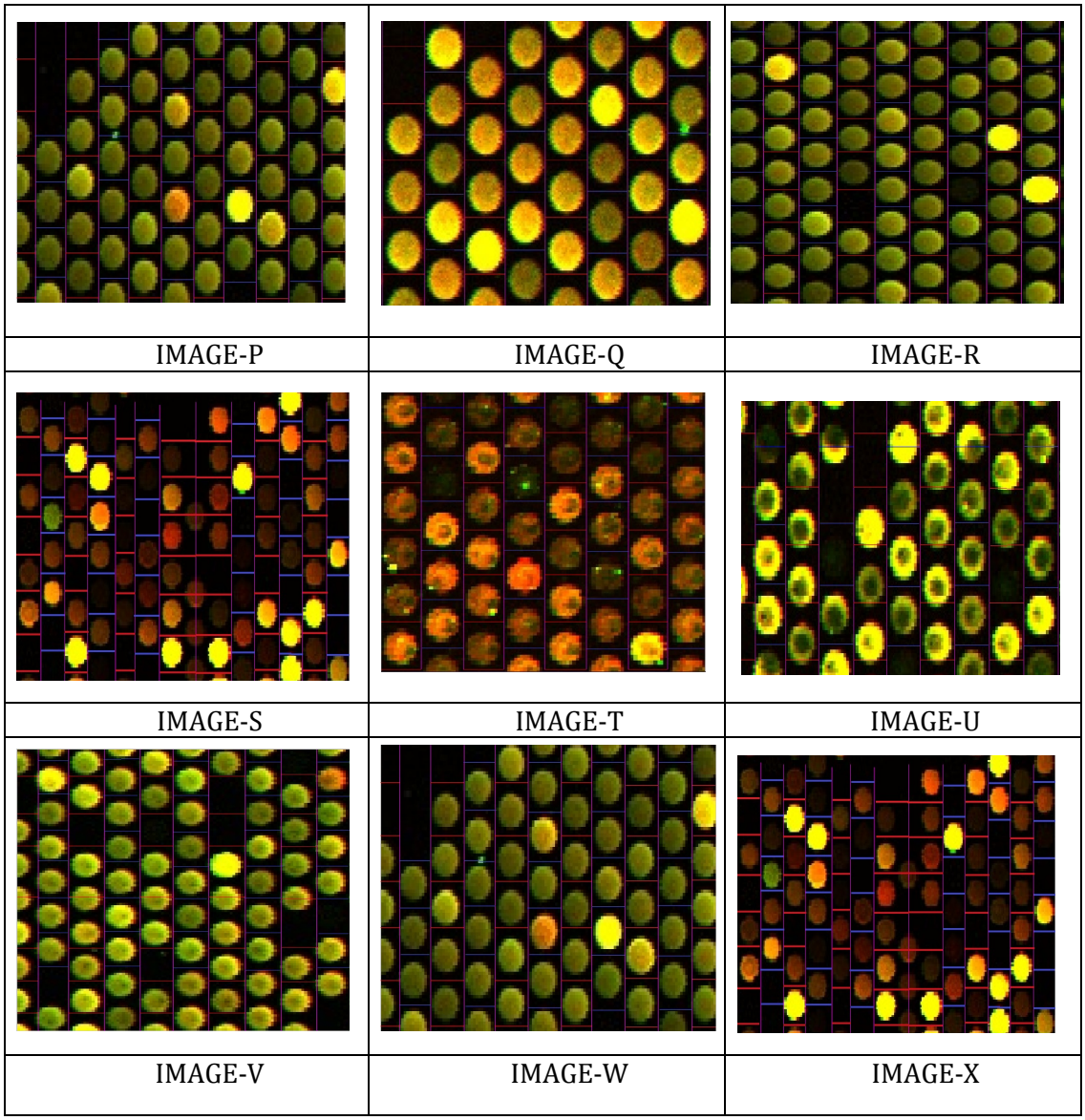


Figure 12: Results of hexagonal gridding using alternate-column

Table 1: Accuracy results for hexagonal spot arrangement (a) individual image gridding accuracy (b) average gridding accuracy based on image categories.

	Method	Diamond-shape	Alternate-column
Category	Image	Accuracy (%)	Accuracy (%)
Best	IMAGE-P	100	100
	IMAGE-Q	100	94
	IMAGE-R	100	100
Typical	IMAGE-S	96	86
	IMAGE-T	92	91
	IMAGE-U	95	100
Worst	IMAGE-V	85	83
	IMAGE-W	86	88
	IMAGE-X	88	87
Average		93.6	92.1

(a)

	Diamond-shape	Alternate-column
Category	Accuracy (%)	Accuracy (%)
Best	100.0	98.0
Typical	94.3	92.3
Worst	86.3	86.0

(b

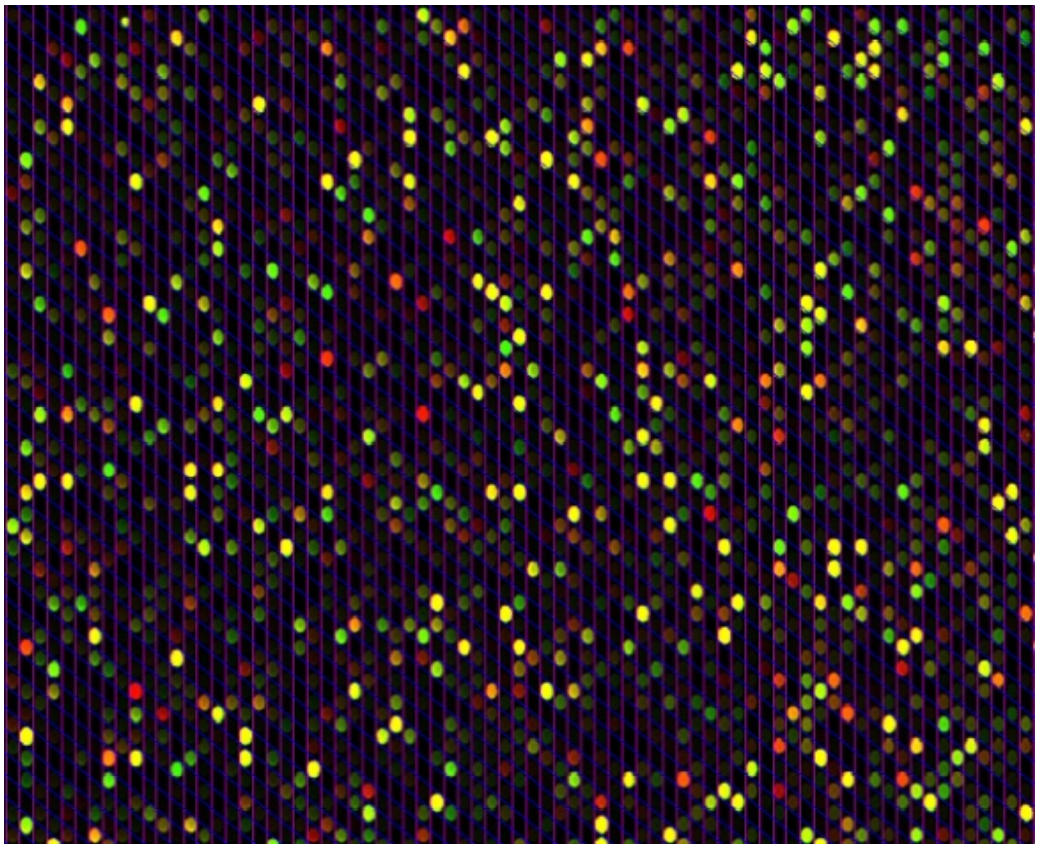


Figure 13: Gridding result for the proposed algorithm using diamond-shape

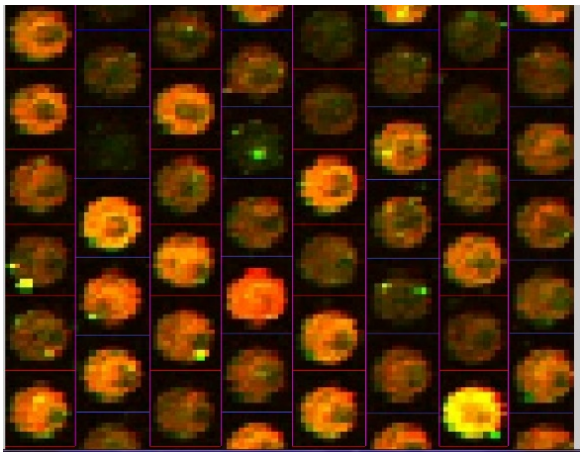


Figure 14: Gridding result for the proposed algorithm using alternate-column

Table 2 and Figure 15 show the computational load for the proposed alternate-column and diamond-shape method. Based on the result, the diamond-shape method consumes lesser computational load compared to the alternate-column method by 68%. Alternate-column consumes three times more computational load because this method performs a horizontal profile for each of the sub-images for each odd column and even column. This contributes toward additional computational time for alternate-columns.

Table 2: Computational load for alternate-column and diamond-shape gridding method

Image	Alternate-column (sec)	Diamond-shape (sec)
IMAGE-P	0.0223	0.0071
IMAGE-Q	0.0234	0.0075
IMAGE-R	0.0218	0.007
IMAGE-S	0.024	0.0066
IMAGE-T	0.0214	0.0068
IMAGE-U	0.0249	0.0073
IMAGE-V	0.0209	0.007
IMAGE-W	0.0202	0.0071
IMAGE-X	0.0177	0.0065
Average	0.021844	0.006989

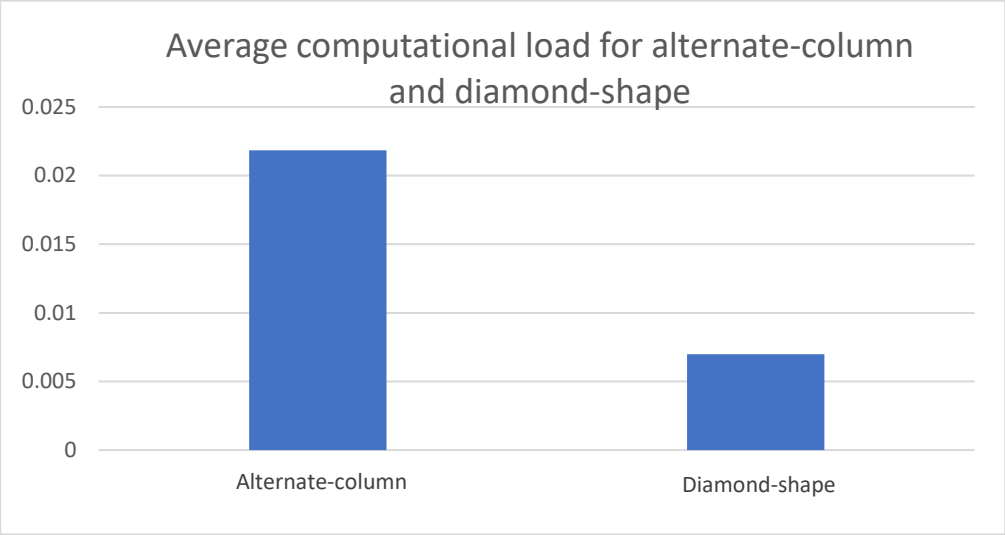


Figure 15: Average computational load for alternate-column and diamond-shape gridding method

Table 3 shows the result comparison for various existing methods as discussed in [16,17], which were used to grid microarray spots arranged in hexagonal and rectangular grids. From the table, our proposed method can achieve up to 100% spot gridding accuracy for hexagonal microarray arrangement as compared to other existing methods. This is achieved within 0.0075 seconds of computational time. This shows that the proposed methods are suitable to be applied to microarray images and produce good results in terms of accuracy and computational load.

Table 3. Results comparison of the proposed gridding methods

Reference/ Dataset	Method description	Image, grid type	Number of spots	Accuracy (%)	Computational time (sec)
SMD [16]	Gridding based on Tophat filtering	Real, Rectangular grid	10294416	98	0.0023
SMD [16]	Wiener filtering	Real, Rectangular grid	10294416	78	0.0034
SMD [16]	Opening and closing filtering	Real, Rectangular grid	10294416	36	0.0103
SMD [17]	Gridding based on support vector machines and genetic algorithms	Real, Rectangular grid	9196	96.4	-
SMD [12]	Voronoi diagrams	Real, Rectangular and hexagonal grids	Various sizes	97.5	-
SMD (present)	Alternate column	Real, Hexagonal grid	42735	94	0.0224
SMD (present)	Diamond shape	Real, Hexagonal grid	42735	100	0.0075

V. CONCLUSION

This paper discusses our proposed methods of gridding for DNA microarray spots arranged in a hexagonal fashion. In this paper, two methods are proposed known as alternate-column and diamond-shape methods. The method is verified using nine benchmark images that represent best, typical and worst case images. Based on our experimental result, our proposed method is able to achieve 93.6% and 92.1% accuracy for diamond-shape and alternate-column, respectively. In terms of computational time, the diamond-shaped method is able to perform the gridding in 0.00699s while the alternate column consumes 0.0218s. This shows that the diamond-shaped method is able to perform the gridding three times faster compared to the alternate-column method. Moving forward, this hexagonal gridding approach could be integrated with more intuitive software, enabling its application across a broad spectrum of images featuring hexagonal arrangements. This method holds the potential for broader adoption of microarray technology, facilitating easier DNA analysis.

REFERENCES

[1] Buckley. M. Duboit. S. Speed. T. & Yang Y, "Comparison of Methods for Image Analysis on cDNA Microarray Data," *Journal of Computational and Graphical Statistics,* vol. 11, pp. 108-136, 2002.

[2] M. Shalon. D. Davis. R. &. Brown. P. Schena, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science,," vol. 270, pp. 467-470, 1995.

[3] S. &. S. Joseph, "CDNA microarray image enhancement for effective gridding of spots,," in *Proc. IEEE Region 10 Conf. (TENCON)*, 2019.

[4] R. Nagarajan, "Intensity-based segmentation of microarray images.," *IEEE Transactions on Medical Imaging,* vol. 22, no. 7, pp. 882-889, 2003.

[5] P. Liu. L. &. Band. M. Bajcsy, "DNA Microarray Image Processing. DNA Array Image Analysis:," Nuts & Bolts (Nuts & Bolts series)., 2007.

[6] E. &. M. D. Zacharia, " An original genetic approach to the fully automatic gridding of microarray images," *IEEE Transactions on Medical Imaging,* vol. 27, no. 6, 2008.

[7] P. Steffy Maria Joseph, "A Fully Automated Gridding Technique for Real Composite cDNA Microarray Images.," *IEEE,* pp. 39605-39622, 2020.

[8] M. M. A. S. Islam Fouad, "Automaticand Accurate Segmentation of Gridded cDNA Microarray Images Using Different Methods," *Advances in Computing,* vol. 4, no. 2, pp. 41-54, 2014.

[9] N. Y. *. Sunil Bhutada, "Opening and closing in morphological image processing," *World Journal of Advanced Research and Reviews,* pp. 687-695, 2022.

[10] M. S. R.M. Farouk, "Microarray spot segmentation algorithm based on integro-differential," *Egyptian informatics Journal,* vol. 20, pp. 173-178, 2019.

[11] I. F. M. M. A. S. Fatma El-Zahraa Labib, "An Efficient Fully Automated Method for Gridding Microarray Images," *American Journal of Biomedical Engineering,* vol. 2, no. 3, pp. 115-119, 2012.

[12] F. K. M. G. T. D. I. F. Nikolaos Giannakeas, "Spot addressing for microarray images structured in hexagonal grids," in *Computer Methods and Programs in Biomedicine*, 2012.

[13] W. W. a. S. H. Steinfath M, "Automated image analysis for array hybridization experiments. Bioinformatics," *oxford,* vol. 17, no. 7, pp. 634-641, 2001.

[14] C. C. G. ,. B. T. a. B. B. Aurel Baloi, "Hexagonal-Grid-Layout Image Segmentation Using Shock Filters: Computational Complexity Case Study for Microarray Image Analysis Related to Machine Learning Approaches," *Sensors,* vol. 23, no. 5, 2023.

[15] Internet, http://smd.stanford.edu.

[16] A. B. J. Y. b. M. Maziidah Mukhtar Ahmad, "Image Gridding Algorithm for DNA Microarray Analyser," in *3rd International Conference on Electronic Design (ICED)*, Phuket, Thailand, 2016.
[17] Bariamis, D.; Maroulis, D.; Iakovidis, D.K. "Unsupervised SVM-based gridding for DNA microarray images". Comput. Med. Imaging Graph, 34, 418–425,2010.

AUTHOR BIOGRAPHY

MAZIIDAH MUKHTAR AHMAD was born in Kedah Malaysia in 1989. She is currently pursuing her final year for her PhD in Microelectronic Engineering in Universiti Malaysia Perlis.UniMAP in 2024. Her current interests also include research in image processing for DNA microarray images.