

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Multivariate Technique for Detecting Variations in High-Dimensional Imagery

RIDWAN A. SANUSI^{1*}, JIMOH OLAWALE AJADI^{2,3*}, SADDAM AKBER ABBASI^{4,5}, TAOFIK O. DAUDA⁶, AND NURUDEEN A. ADEGOKE⁷

¹Interdisciplinary Research Center for Smart Mobility and Logistics, King Fahd University of Petroleum Minerals, Dhahran, 31261 Saudi Arabia. (e-mail: amhigher2010@yahoo.com)

²Center for Refining Advanced Chemicals, King Fahd University of Petroleum Minerals, Dhahran, 31261 Saudi Arabia

³Department of Mathematics, King Fahd University of Petroleum Minerals, Dhahran, 31261 Saudi Arabia. (e-mail: jimoh.ajadi@kfupm.edu.sa)

⁴Statistics Program, Department of Mathematics, Statistics, and Physics, College of Arts and Sciences, Qatar University, 2713, Doha, Qatar.(e-mail: sabbasi@qu.edu.qa)

⁵Statistical Consulting Unit, College of Arts and Sciences, Qatar University, 2713, Doha, Qatar

⁶Institute of Agriculture Research and Training, Obafemi Awolowo University, Moor Plantation Ibadan, Nigeria. (e-mail: todauda.iart@gmail.com)

⁷Melanoma Institute Australia, The University of Sydney. (e-mail: nurudeen.adegoke@sydney.edu.au)

*: Contributed Equally

Corresponding author: Saddam Akber Abbasi (e-mail: sabbasi@qu.edu.qa).

ABSTRACT The field of immunology requires refined techniques to identify detailed cellular variance in high-dimensional images. Current methods mainly capture general immune cell proportion variations and often overlook specific deviations in individual patient samples from group baseline. We introduce a simple technique that integrates the Multivariate Shewhart Control Chart (MSCC) with random projection (RP) methods, specifically designed to identify changes in immune cell composition in high-dimensional images. Uniquely, our method provides deeper insights into individual patient samples, allowing for a clearer understanding of group differences. We assess the efficacy of MSCC across various RPs: Achlioptas (AP), plus-minus one (PM), Li, and normal projections (NP), considering shift size, dimension reduction, and image dimensions. Simulations reveal variable detection performances across RPs, with PM outperforming and Li lagging. Practical tests using single-cell images of basophils (BAS) and promyelocytes (PMO) emphasise their utility for individualised detection. Our approach elevates high-dimensional image data analysis, particularly for identifying shifts in immune cell composition. This breakthrough potentially transforms healthcare practitioners' cellular interpretation of the immune landscape, promoting personalised patient care, and reshaping the discernment of diverse patient immune cell samples.

INDEX TERMS Dimensionality reduction, High-dimension data, Image monitoring, Multivariate Shewhart control chart, Quality control in healthcare, Random projection methods.

I. INTRODUCTION

THE biomedical and biomedical engineering fields have experienced significant growth owing to advancements in technology that have increased the availability and analysis of high-dimensional data. Investigating changes in immune cell composition using multiplexed immunohistochemical (mIHC) images is a particularly dynamic area in these fields [1]. Understanding the immune landscape of various diseases, particularly cancer, has the potential to provide valuable insights into the transformation of diagnostic and therapeutic strategies to enhance patient outcomes [2], [3]. Traditional statistical tests, such as t-tests (for means) or the Wilcoxon Rank Sum test (for medians), can be used to compare immune cell compositions between groups. However,

these tests are limited in their ability to provide comprehensive insights into variances (or differences) within each patient sample. This limitation can restrict the interpretation of cellular heterogeneity and immune response dynamics, making it challenging to make precise clinical decisions [4]. This highlights the need for more advanced and sophisticated techniques, such as the application of multivariate control charts, specifically Multivariate Shewhart control charts (MSCC), implemented under random projection (RP). This approach has the potential to detect and quantify changes in immune cell composition in individual samples, providing more granular insights that can inform clinical decision making [5], [6].

MSCC, a well-known and highly regarded tool for quality

and statistical process control, is widely recognised for its ability to monitor and detect significant changes in the data. However, its utility in high-dimensional data, such as mIHC images, has been limited because of the challenges of computational complexity and difficulty in distinguishing genuine changes from the complexity of the data [6]. Dimensionality reduction methods, specifically the RP, offer solutions to these limitations. RP is an effective technique for reducing the dimensionality of a dataset while preserving its underlying structure, making it simpler and easier to analyse [7]. In the context of MSCC, RP can enhance the sensitivity by focusing on the most relevant dimensions, thereby improving the detection of changes [8]. For instance, by reducing the dimensionality of an mIHC image, we can efficiently track changes in the immune landscape of each patient without losing critical information regarding the different immune cell compositions.

To enhance the application of RP in the context of MSCC, this study examines four distinct RP methods: Achlioptas, Plus-Minus One, Li, and Normal projections. Each method has its own unique characteristics. For instance, Achlioptas projection (AP) [9] can reduce computational demands by introducing sparsity into the projected data. Plus-Minus One projection (PM) provides a computationally efficient approach owing to its simple binary nature. Li projection (Li) [10], which focuses on preserving pairwise distances, ensures an accurate representation of the original data. Finally, normal projection is robust to variations in data [10]. Our objectives are to explore how each method affects the performance of MSCC and to understand the implications of their unique characteristics in detecting changes in immune composition.

The performance of the proposed RP-based MSCC approach is highly dependent on its ability to accurately and rapidly detect changes in high-dimensional image data streams. The assessment of the performance of the approach is primarily based on run length properties, specifically the average run length (ARL) and the standard deviation of the run length (SDRL) [11]. These metrics serve as robust statistical measures for comprehensively evaluating the sensitivity and specificity of a chart. Specifically, the ARL indicates the chart's sensitivity, which is the average number of observations required to detect a shift or change. SDRL indicates variability in run length. Therefore, a balance between these two measures is crucial for effective monitoring.

In this study, we propose an innovative method for determining changes in high-dimensional immune cell data that can serve as a complementary approach to conventional methods such as t-tests or the Wilcoxon Rank Sum test. Our aim is to contribute to the development of more efficient and effective monitoring tools for the immune microenvironment by carefully observing changes in immune cell composition within individual patients and across different patient groups. This approach has the potential to advance our understanding of cellular interactions in the immune system and lead to improved prognostic and therapeutic strategies for healthcare

[12]. Given its significant implications for healthcare and the potential for substantial improvements in patient care, this study has both practical and theoretical importance. Theoretically, this study can enhance our understanding of the application of control charts to high-dimensional biomedical data. Practically, this approach can guide healthcare practitioners in making data-driven decisions that can significantly improve patient health outcomes.

The remainder of this paper is organised as follows. Section II discusses different RP methods, their dimensionality determination, and the process of obtaining reduced spaces using these methods. Section III presents a strategy for incorporating RP methods into the MSCC. Section IV presents extensive simulation studies. Section V describes the application of MSCC under RPs to monitor changes in immune cell composition within image datasets. Section VI discusses the findings, broader significance, potential applications, and limitations of the study, and suggests avenues for future research. Finally, Section VII concludes the paper by summarising key insights of the study.

II. OVERVIEW OF RANDOM PROJECTIONS

The increasing prevalence of high-dimensional data in various fields presents a significant challenge to traditional data analysis methods due to the "curse of dimensionality" [13]. However, Random Projection (RP) has gained popularity as a powerful dimensionality reduction method, chiefly because of its simplicity, computational efficiency, and robust theoretical foundations [7]. The core principle of RP involves transforming high-dimensional data into a lower-dimensional space using a random matrix, while preserving pairwise distances and the underlying data structure [14]. Its strength lies in its universal applicability, as it can be applied across various data types and domains without relying on data-specific assumptions. In addition, the generation of random matrices can be tailored to achieve different trade-offs, thereby providing flexibility for various applications [10], [15].

At its core, RP approximates the distance between two vectors, that is \mathbf{x} and \mathbf{z} , in a p -dimensional space as follows:

$$D^2(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|^2 = (\mathbf{x} - \mathbf{z})^T (\mathbf{x} - \mathbf{z}), \quad \mathbf{x}, \mathbf{z} \in \mathbb{R}^p,$$

where $D^2(\cdot)$ is the squared distance. The estimate of the sample squared distance is given as:

$$\hat{D}^2(\mathbf{x}, \mathbf{z}) = \frac{1}{d} \sum_{j=1}^d (r_j(\mathbf{x} - \mathbf{z}))^2 = \frac{1}{d} \|\mathbf{R}\mathbf{x} - \mathbf{R}\mathbf{z}\|^2,$$

where r_j ($j = 1, 2, \dots, d$) denotes the i -th row of \mathbf{R} (representing a unique projection), $d \ll p$ is the reduced dimensionality, and \mathbf{R} is a $d \times p$ RP matrix. For vectors $\mathbf{x}, \mathbf{z} \in \mathbb{R}^p$:

$$\begin{aligned} E(\hat{D}^2) &= D^2, \\ \text{var}(\hat{D}^2) &= \frac{2}{d} D^4, \\ \frac{d\hat{D}^2}{D^2} &\sim \chi_d^2. \end{aligned}$$

These suggest that:

$$\Pr \left(\frac{|\hat{D}^2 - D^2|}{D^2} \geq \epsilon \right) \leq 2 \exp \left(-\frac{d}{4}\epsilon^2 + \frac{d}{6}\epsilon^3 \right),$$

with $\epsilon \in (0, 1)$ [14], [16], [17]. Additionally, given the linearity of RP represented by $\mathbf{R} \in \mathbb{R}^{d \times p}$:

$$\mathbf{R} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{R} \mathbf{z}_i, \quad (1)$$

it is clear that any variation in the projected vectors $\mathbf{R} \mathbf{z}_1, \dots, \mathbf{R} \mathbf{z}_n$ mirrors the changes in the original average $\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i$. This property extends to a broader "mean invariance under the linear RP" characteristic, forming the foundation of our proposed method. When $\mathbf{X}_i, i = 1, \dots, n$ are i.i.d. vectors with a finite first moment:

$$E \left\{ \mathbf{R} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right) | \mathbf{R} \right\} = \mathbf{R} E(\mathbf{X}),$$

where $E(\cdot | \mathbf{R})$ represents the conditional expectation given \mathbf{R} .

A. RP PRESERVATION LEMMA

The Johnson-Lindenstrauss (JL) lemma has significantly contributed to the development of the RP. This lemma asserts that points in high-dimensional spaces can be effectively projected onto a lower-dimensional space, while maintaining their pairwise distances [14], [18].

JL Lemma: Given a set of points \mathbf{X} in \mathbb{R}^p and an error tolerance $0 < \epsilon < 1$, there exists a linear map $f: \mathbb{R}^p \rightarrow \mathbb{R}^d$ ensuring

$$(1 - \epsilon) \|\mathbf{x} - \mathbf{z}\|_2^2 \leq \|f(\mathbf{x}) - f(\mathbf{z})\|_2^2 \leq (1 + \epsilon) \|\mathbf{x} - \mathbf{z}\|_2^2, \quad (2)$$

This lemma guarantees near-perfect distance preservation under map f and implies that a reduced-dimensional representation of the data can be maintained with an error tolerance depending on ϵ [18]. The Euclidean distance should preferably remain unchanged after the representation

$$\|\mathbf{R} \cdot \mathbf{x} - \mathbf{R} \cdot \mathbf{z}\|^2 = \|\mathbf{x} - \mathbf{z}\|^2, \quad (3)$$

The JL lemma provides a theoretical foundation, but does not provide a method for constructing the map f . This work explores several methods, including Achlioptas, Plus-Minus One, Li, and Normal projections, to address this gap and offer efficient ways to achieve dimensionality reduction. These methods are appealing because of their simplicity, efficiency, and independence of specific data distributions.

B. DIFFERENT RP TYPES

1) Achlioptas Projection

AP [9] introduces sparsity into an RP matrix, resulting in a method with remarkable computational efficiency, particularly when processing high-dimensional datasets. This

method utilises sparse matrices in RPs, which can significantly accelerate computations and conserve storage space [19]. In the AP, the RP matrix, denoted by \mathbf{R} , is populated such that each entry r_{jl} $l = 1, 2, \dots, p$, is computed based on a specific distribution:

$$r_{jl} = \sqrt{3} \times \begin{cases} +1 & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -1 & \text{with probability } \frac{1}{6}, \end{cases} \quad (4)$$

This distribution ensures that two-thirds of the entries in matrix \mathbf{R} are zero, thus making it a sparse matrix. Multiplication by $\sqrt{3}$ preserves the pairwise Euclidean distances and maintains the structural integrity of the data in a lower-dimensional space [15]. The unique sparse structure introduced by AP makes it an attractive choice for high-dimensional data reduction in situations where computational resources are limited. Despite its simplicity, it can effectively capture the key features of data, facilitating efficient processing and analysis [19].

2) Plus-Minus One projection

PM is a method for generating a RP matrix that distinguishes itself from other methods by not having zero entries, also known as full density. This lack of zeros enables the better preservation of distances and structures in the data [10]. The methodology behind PM is simple and powerful. The full-density RP matrix \mathbf{R} is populated by computing each entry r_{jl} independently following a specific distribution:

$$r_{jl} = \begin{cases} +1 & \text{with probability } \frac{1}{2}, \\ -1 & \text{with probability } \frac{1}{2}. \end{cases} \quad (5)$$

This uniform distribution allows for straightforward and fast computation, making PM a computationally less intensive method. The assigned values, +1 or -1, are distributed randomly throughout the matrix, ensuring that the projected data maintains a strong representation of the original high-dimensional space [20]. The simplicity and computational efficiency of PM make it an attractive choice for handling large-scale high-dimensional data. They are particularly useful in applications in which the preservation of the original data structure is of great importance. Furthermore, because there are no zero entries, PM can result in more diverse projections, which can be advantageous in certain scenarios [7].

3) Li Projection

Li is an alternative approach to creating sparse RP matrices [10]. As a dimensionality reduction method, Li was designed to strike a balance between computational efficiency owing to the sparsity of the projection matrix and the preservation of the original data structure. Sparsity in the Li matrix arises from half of the entries in the projection matrix, which are set to zero. This distribution is obtained by ensuring that each

entry r_{jl} in the RP matrix \mathbf{R} is computed using the following formula:

$$r_{jl} = \begin{cases} +1 & \text{with probability } \frac{1}{2s}, \\ 0 & \text{with probability } 1 - \frac{1}{s}, \\ -1 & \text{with probability } \frac{1}{2s}, \end{cases} \quad (6)$$

where $s = \lceil \sqrt{p} \rceil$ denotes scaling factor. This projection introduces sparsity into the projection matrix with three potential values (+1, 0, -1). In a sparse setup, multiplications with zero entries do not contribute to the final results, thereby enhancing the computational efficiency of high-dimensional data [21]. The use of +1, -1, and 0 assignments simplifies the process, removing the need to scale non-zero entries, which further reduces the computational overhead. Li demonstrated its effectiveness in a range of applications, particularly those requiring both dimensionality reduction and computational efficiency. By striking a balance between these two critical factors, Li serves as an efficient tool for managing high-dimensional data [22], [23].

4) Normal Projection

NP (also known as Gaussian Projection) utilises the entries of an RP matrix drawn independently from a normal distribution [13]. This method has a solid statistical foundation because it relies on the isotropic properties of the normal distribution, which remains unchanged under any rotation of the coordinate system. This isotropic property is particularly beneficial for preserving the structure and relationships within high-dimensional data during dimensionality reduction. The RP matrix \mathbf{R} in NP is populated such that each entry r_{jl} is independently drawn from a standard normal distribution, as follows:

$$r_{jl} \sim \mathcal{N}(0, 1). \quad (7)$$

This normal distribution ensures that the entries in the matrix are balanced, with most of the entries being small (close to zero), but with a chance for larger values as well. This balance can provide a fair representation of the original data when projected onto a lower-dimensional space [7].

C. DETERMINING THE DIMENSION OF RPS

Appropriate projection dimensionality in RP methods is crucial because it affects data quality and computational efficiency. The goal is to balance the data structure while ensuring feasible computations. The JL lemma provides guidance on this matter [14], allowing for significant dimension reduction while preserving pairwise distances to a distortion factor ϵ [9]. This leads to a more efficient data processing and analysis. The JL lemma formula estimates the appropriate dimension d for a dataset with cardinality n , given by $\lceil 4 \cdot \frac{\log(n)}{\epsilon^2} \rceil$. The ceiling function ensures that d is an integer. d derived from the JL lemma ensures the preservation of pairwise distances in the original high-dimensional data in the lower-dimensional space within the factor ϵ [9]. This formula provides practical guidance for setting the projection

dimension, balancing data structure preservation, and computational efficiency [21] and is dependent on the number of data points and the desired precision (ϵ) rather than the original data dimensionality. This enhances computational feasibility, particularly for high-dimensional datasets [24].

D. OBTAINING THE REDUCED SPACE THROUGH RP

The fundamental operation performed by RP methods is matrix multiplication between the original data matrix and the projection matrix. This operation reduces the dimensionality of the data while preserving their essential properties. Let \mathbf{X} be an $n \times p$ matrix, where n represents the number of data points and p is the original dimensionality. The RP matrix \mathbf{R} , once generated, is kept constant. The projected data matrix \mathbf{Y} with dimensions $n \times d$ can be obtained by multiplying \mathbf{X} by \mathbf{R}' , where \mathbf{R}' is the transpose of \mathbf{R} .

III. METHODOLOGY FOR INTEGRATING RPS INTO MSCC

We focus on examining the changes in multichannel images, especially RGB images, represented as an $n \times p \times 3$ tensor \mathbf{X} , where p signifies the number of features or quality characteristics (e.g. colour, texture, and size), n represents the number of instances in which each feature has been measured or observed, and the third dimension corresponds to the RGB channels, encapsulating the intricate color information inherent in the images. The multichannel nature of the data provides a rich source of information that is critical for various image-analysis tasks. Each channel in the RGB image channel can be analysed independently or jointly to gain comprehensive insight into the underlying structure and attributes of the images in question. This RGB-focused approach facilitates a more nuanced understanding, enabling thorough exploration of the multivariate data inherent in multichannel imagery. This approach is particularly powerful as it facilitates the disentanglement of complex patterns and relationships embedded in the RGB channels. This method is suitable for various practical applications in which distinguishing and quantifying subtle changes across RGB channels is essential. Specifically, it provides a robust framework for tasks such as image classification, object detection, and quality assessment, where the ability to discern and measure variations in color, texture, and size contributes significantly to the overall efficacy of the analysis.

The initial stage of our methodology involves image segmentation to isolate regions of interest (ROIs) from the surrounding regions. This process begins with the conversion of images to gray scale, followed by thresholding using the Otsu method. This technique effectively delineates cellular regions from the background by identifying an optimal threshold value that maximises the variance between two classes of pixels. The segmentation tasks are conducted using the EBImage package [25] in the R 4.3.0 program [26], resulting in binary images in which the regions of interest are distinctly isolated from the background. Interested readers

can see relevant literature on isolating and extracting ROIs using image segmentation [27]–[30]

Upon successful segmentation, the subsequent phase entails the extraction of features from the segmented ROIs. Recent advancements in the field underscore the significance of feature extraction for enhancing the discriminatory power and interpretability of image analysis methods [31]–[34]. Our matrix representation \mathbf{X} undergoes feature extraction to populate it with meaningful attributes that encapsulate image composition. These features include colour, texture, and size, which are pivotal in the MSCC framework. To extract the colour features, we operate within the RGB colour space, widely acknowledged for its representation of colour information in digital images. In particular, we compute the grayscale of the image, distilling the intricate colour details into a single value, aiding further analysis within the MSCC framework. This conversion not only simplifies the subsequent analytical processes but also preserves the essential colour information required for our methodology. This step of converting the image into a single channel underscores the adaptability of the methodology for handling multichannel image data and transforming them into a structured format that is conducive to subsequent analytical processes within the MSCC framework. The MSCC framework is an established method in image analysis, recognized for its efficacy in identifying patterns and deviations within multichannel images [35].

The methodology employed in this study comprises three primary stages: defining the MSCC limits, controlling the operation of the MSCC, and evaluating the performance of the MSCC using the run length distribution of the control chart. Our methodology utilises AP, PM, Li, and NP reliability programs to reduce the dimensions of the original image, represented as the reduced image $\mathbf{Y}_{n \times d}$. The reduced dimension is determined by the size of the original data and the desired precision level, which dictates the acceptable level of distortion during the reduction process. Subsequently, the reduced image $\mathbf{Y}_{n \times d}$ is fed into the MSCC. A notable feature of MSCC is its emphasis on current observations, which renders the chart highly sensitive to changes. This sensitivity enables the immediate detection of significant changes.

For the k th image, the MSCC signals a change when the following equation is satisfied:

$$T_k^2 = n(\bar{\mathbf{Y}}_k - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{Y}}_k - \boldsymbol{\mu}_0) > h, \quad (8)$$

where,

$$\bar{\mathbf{Y}}_k = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_{ki} \quad (9)$$

where $\bar{\mathbf{Y}}_k$ is the average of the reduced-dimension representation of the k th image, $\boldsymbol{\mu}_0$ is the in-control mean vector, and $\boldsymbol{\Sigma}_0$ is the in-control covariance matrix. The upper control limit (UCL), h , is selected based on the desired in-control performance measure, such as the desired value of the in-control ARL (ARL_0). If the in-control parameters ($\boldsymbol{\mu}_0$ and

$\boldsymbol{\Sigma}_0$) are unknown (Case U), they can be estimated from reliable historical data [36], [37].

A. ESTIMATION OF $\boldsymbol{\mu}_0$ AND $\boldsymbol{\Sigma}_0$ PARAMETERS

The RP-based MSCC method involves two phases: Phase I, also known as the "image capability analysis phase" or the "baseline phase," and Phase II, or the "control phase." During Phase I, historical imaging data are collected and analyzed to estimate the relevant parameters of immune cell composition and understand how they behave at baseline [38], [39]. This foundational understanding is then used to construct the control charts. Any unique variations in immune cell composition identified during Phase I require detailed examination. Once the parameters are established and the control limits are set, the transitions are monitored until Phase II. During Phase II, real-time monitoring of immune cell compositions in imaging datasets is performed using the control limits set in Phase I. Decisions are informed by signals from the control chart, such as when a point falls outside the limit, indicating potentially significant changes in immune cell composition that may require further scrutiny [40].

In our methodology, the initial step of Phase I involves estimating the parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ for the T^2 statistic (as shown in Equation 13). We follow a straightforward two-step approach to obtain estimates for each image. First, we apply a specific RP method to the k th image, resulting in \mathbf{Y}_{ki} , where \mathbf{Y}_{ki} is the i th sample for the k th image, $i = 1 \dots n$, $k = 1 \dots n_{\text{phase1}}$, and n_{phase1} is the number of images in Phase I. Then, we compute $\bar{\mathbf{Y}}_k$ (Equation 9). Subsequently, the estimated covariance matrix of the k th projected image is computed as follows:

$$\hat{\mathbf{S}}_k = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k)(\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k)' \quad (10)$$

Next, $\boldsymbol{\mu}_0$ is estimated as the average of $\bar{\mathbf{Y}}_k$, i.e.,

$$\hat{\boldsymbol{\mu}}_0 = \frac{1}{n_{\text{phase1}}} \sum_{k=1}^{n_{\text{phase1}}} \bar{\mathbf{Y}}_k \quad (11)$$

The covariance matrix parameter, $\hat{\boldsymbol{\Sigma}}_0$, is estimated as the average of all covariance matrices across all images in the Phase I dataset as follows:

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n_{\text{phase1}}} \sum_{k=1}^{n_{\text{phase1}}} \hat{\mathbf{S}}_k \quad (12)$$

The proposed RP-based Hotelling's T^2 statistic based on $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$ used for monitoring in Phase II is given as

$$\hat{T}_k^2 = n * (\bar{\mathbf{Y}}_k - \hat{\boldsymbol{\mu}}_0)' \hat{\boldsymbol{\Sigma}}_0^{-1} (\bar{\mathbf{Y}}_k - \hat{\boldsymbol{\mu}}_0). \quad (13)$$

B. ESTABLISHING SHEWHART CONTROL LIMIT

The distribution of MSCC based on RP is unknown. Consequently, traditional parametric methods for setting control

limits are not feasible in this context. To address this challenge, an empirical limit is utilised. This method is non-parametric and does not require any assumptions regarding the underlying distribution of the Shewhart statistics. Consequently, it is highly flexible and can be applied to a wide range of scenarios, making it a valuable tool for quality control. To detect changes in the immune composition, the UCL of the chart is determined using the empirical distribution of \hat{T}^2 from the n_{phase1} images gathered during Phase I. Specifically, the UCL is chosen by selecting the $1-\alpha$ quantile of the \hat{T}^2 values from the Phase I data as the UCL:

$$\text{UCL} = Q_{1-\alpha}(\hat{T}_1^2, \hat{T}_2^2, \dots, \hat{T}_{n_{\text{phase1}}}^2), \quad (14)$$

where $Q_{1-\alpha}$ denotes the $1 - \alpha$ quantile of the Shewhart statistic and $\alpha = \frac{1}{\text{ARL}_0}$.

IV. SIMULATION STUDY AND RESULTS

A. SIMULATION STUDY

In this study, we evaluate the performance of the MSCC with respect to the detection of shifts in the original image data, which represents an out-of-control state in the image datasets. To achieve this, we introduce shifts in the original image data and calculate the ARL corresponding to each shift, providing a comprehensive assessment of the chart's performance under various shift scenarios. Additionally, we employ SDRL method to compare the performance of the MSCC across different shift scenarios. To achieve this, we carefully select a range of relevant and potentially efficient RP methods and evaluate their effectiveness when used in conjunction with the MSCC for high-dimensional image data. Our goal is to identify the most effective method for detecting shifts in image data that demonstrates the best performance.

We employ Gaussian noise to simulate shifts in the original image data by formulating the following equation

$$p_{ij,\text{new}} = p_{ij,\text{old}} + N(\delta, s_{0,\text{change}}^2) \quad (15)$$

where $p_{ij,\text{old}}$ represents the original pixel value at the i -th row and j -th column, $N(\delta, s_{0,\text{change}}^2)$ is a random number drawn from a Gaussian distribution with mean δ and variance $s_{0,\text{change}}^2$, and $p_{ij,\text{new}}$ is the new pixel value after the noise has been added. Parameter δ signifies the desired shift in pixel intensity, and $s_{0,\text{change}}^2$ represents the variability in the intensity shift across the image. Gaussian noise is utilised because it closely mimics real-world scenarios. A comprehensive evaluation of the performance of the RP methods is achieved by varying the following parameters: δ , RP methods, image size, and ϵ .

We conduct a comprehensive analysis by examining various values of δ (i.e., 0.025, 0.05, 0.075, 0.1, 0.2, 0.4, 0.6, 0.8, and 1) to investigate the impact of different shifts on the image properties. We employ the following RP methods: NP, Li, PM, and AP. Our aim is to identify the most effective approach for the MSCC by testing these different methods. We select two different image sizes, 100×100 and 142×142 pixels, to explore the influence of data dimensionality on the performance of the MSCC. To evaluate the performance of

the MSCC under varying dimensionality, we vary ϵ such that $d = 2, 5, 10, 15, 20$ and 30 . We set the simulation size and Phase I data size to $50,000$ and $1,000,000$, respectively, to ensure high performance of the MSCC. To facilitate a standardised evaluation, we set the ARL to a fixed target of 200 in accordance with previous studies [41], [42]. This allows a consistent benchmark to measure and understand the performance of the chart in a controlled setting.

B. SIMULATIONS RESULTS

The simulation results for each RP with varying values of d are presented in Tables 1-6. The findings from our simulation studies provide valuable insights into how various parameters and adjustments affect the ARL and SDRL of the MSCC. The following discussion details the effects of each factor.

1) Effect of change in δ

The performances of the RPs are comparable when δ is zero. However, as δ increases, the performance of RPs diverges. For example, at $\delta = 0.025$, an image size of 100×100 , and $d = 2$ (Table 1), the ARL (SDRL) values for Li, AP, NP, and PM are 163.47 (163.4), 162.89 (163.89), 155.97 (154.78), and 154.66 (153.86), respectively. As the magnitude of the change increases, the methods detect it more quickly, and the ARL values consistently decreasing. This is intuitive, because larger changes are easier to detect. With a larger shift of $\delta = 0.075$, these values shift to 112.97 (112.07), 103.41 (102.33), 100.12 (100.13), and 96.02 (95.72). With the maximum shift of $\delta = 3$, ARL (SDRL) values are 1.82 (1.23), 1.79 (1.18), 1.79 (1.2), and 1.76 (1.15). This performance is consistent with the other values of image size and d (Tables 1-6).

2) Comparison of RP

At $\delta = 0$ (no change), all RP methods consistently yield ARL values of approximately 200 for both image dimensions, indicating a uniform performance in the absence of any change. However, for $\delta > 0$, all the methods exhibit varying levels of performance. Specifically, the PM method for a 100×100 image size tends to have the lowest ARL (SDRL) for most values of δ (Tables 1-6). For instance, at $\delta = 0.1$, the ARL (SDRL) values for Li, AP, NP and PM methods are 93.08 (92.9), 83.88 (83.62), 81.21 (81.02), and 76.91 (75.7), respectively, for $d = 2$ (Table 1). These values change to 110.87 (110.56), 99.35 (98.77), 103.05 (102.65), and 97.82 (97.81), respectively, for $d = 10$ (Table 3); and further to 116.62 (117.25), 107.21 (106.45), 110.48 (109.2), and 104.28 (104.07), respectively, for $d = 20$ (Table 5). The exceptional efficiency of the PM and Li's lagging performance provide valuable insights into the optimal method selection based on the given parameters. However, for a larger 142×142 image size, the differences in method performance become less distinct, suggesting the influence of image size on method differentiation.

3) Effect of change in image size

When the image size increases from 100×100 to 142×142 , there is a slight increase in the ARL values, particularly for higher values of δ and small values of d . This increase in ARL values for larger images might indicate that the monitoring process is slower in detecting changes in larger images. For instance, when $\delta = 0.025$ and $d = 2$, the ARL (SDRL) values for images of 100×100 and 142×142 are 154.66 (153.86) and 161.63 (161.23), respectively. Similarly, when $\delta = 0.025$ and $d = 5$ (Table 2), the ARL (SDRL) values are 161.74 (161.24) and 163.77 (163.58), respectively.

4) Effect of reduced dimension (d)

As shown in Tables 1–6, as the value of d increases, there is a general trend of an increase in the ARL values. For instance, when $\delta = 0.2$ and the image size is 100×100 pixels, the ARL (SDRL) values for PM under $d = 2, 5, 10, 15, 20, 30$ are 33.68 (33.2), 42.91 (42.8), 50.33 (50.11), 55.04 (54.52), 57.1 (56.55) and 60.87 (60.33), respectively. Similarly, the ARL (SDRL) values for Li are 45.13 (44.38), 55.55 (55.37), 62.51 (61.52), 63.53 (63.31), 68.51 (67.64), and 71.26 (70.43), respectively, as d increases. This indicates that increasing the value of d results in slower shift detection.

V. REAL-LIFE EXAMPLE

Monitoring changes in the composition of immune cells, particularly basophils (BAS) and promyelocytes (PMO), in bone marrow samples can provide crucial insights into the healthcare field. Basophils play a crucial role in immune responses and are associated with allergies and asthma [43]. Abnormal levels of BAS cells indicate allergic reactions or autoimmune disease [44]. In contrast, PMO cells are integral to the development of granulocytes, a type of white blood cells [45]. Any abnormality in PMO cell levels can be a warning sign for acute promyelocytic leukaemia (APL), a type of cancer [46]. The correlation between the BAS and PMO cells underscores the significance of consistent observations. Therefore, monitoring these cells may aid early diagnostic procedures and empower clinicians with timely intervention tools [47].

In our analysis, we utilise the capabilities of RP-based MSCC to monitor multivariate data by focusing on BAS and PMO cell images (Figure 1). The cell image data, sourced from the Munich Acute Myeloid Leukemia (AML) Morphology Dataset [48], comprises 18,365 expert-labeled single-cell images obtained from peripheral blood smears of 200 participants, including AML patients and non-malignant controls. Image acquisition used a M8 digital microscope/scanner at 100-fold optical magnification and oil immersion. The dataset, meticulously annotated by trained experts, has been utilized for training a convolutional neural network for single-cell morphology classification [49]. This pathology dataset provides a rich source for validating the RP-based MSCC method on real-world, high-dimensional image data scenarios. More details on the dataset are available in the literature [48]–[50].

To monitor changes in the BAS and PMO cell morphology [48], we employ 79 BAS and 70 PMO images, each sized at 400×400 pixels, comprising three distinct channels. We conduct extensive image analysis comprising segmentation and feature extraction, to better understand the composition of the images. The segmentation process, which involves gray scale conversion and Otsu's thresholding method, isolates the regions of interest (ROIs) from the background. After segmentation, we extract features such as colour, texture, and size from the ROIs by using the EImage package [25] in R 4.3.0 [26]. While these are significant features for distinguishing cells, other features such as cell morphology, granularity, nucleus-to-cytoplasm ratio, and specific cellular markers expressed on the cell surface are additional features that play a crucial role in distinguishing different cell types. However, this study only focuses on the significant features, i.e., colour, texture, and size values.

Given the inherent complexities of high-dimensional datasets, it is crucial to employ dimensionality reduction techniques to simplify the analysis. By utilising four different RP methods, namely, Li, AP, NP, and PM, we effectively reduce the dimensionality of each image to 400×20 pixels. Two distinct phases are employed in the pipeline. Phase I focuses on BAS cell images and aims to establish a reliable control limit that characterizes the baseline samples. This foundational baseline, derived from Phase I, serves as a vital guide for our Phase II investigations, which delve deeply into the PMO cell images. In the latter phase, we compare their T^2 values against the control limit established in Phase I to detect any deviations from the established baseline. The T^2 values play a pivotal role in discerning differences between the comparing cells. They represent the degree of deviation of multivariate data points from the mean and are instrumental in identifying patterns or anomalies [17]. Also, they are considered indicative of significant variations in the characteristics of the BAS and PMO cells. The conclusive nature of these values is supported by their statistical significance, demonstrating that the observed differences are beyond what might be expected by random chance.

Phase I analysis reveals that all the projection methods identified four samples outside the UCL, suggesting the presence of inherent anomalies or distinct patterns in these samples (Figure 2). This detection across multiple projection methodologies indicates that these samples represent genuine deviations and not random statistical outliers. To mitigate the possible influence of such outliers on subsequent analyses, a rigorous pre-processing step was undertaken, resulting in the removal of these points. Following this cleanup, no samples exceed the UCL, emphasising the crucial role of data pre-processing in establishing a reliable and representative baseline that is essential for the subsequent phases of the study (Figure 3).

The results of Phase II (Figure 4), which are illustrated by the number of flagged samples for each RP method, provide valuable information. The differences in the number of detected samples - 37 by PM, 25 by AP, 27 by NP, and

26 by Li - highlight the varying sensitivity of each method in detecting changes within the immune cell compositions. Specifically, a higher number of detected samples indicates an increased ability to identify subtle alterations. Among these, the PM method demonstrates superior performance in detecting variations during Phase II, suggesting its superior effectiveness in high-dimensional data analysis.

However, it is crucial to understand that these variations might not be inherently problematic but could be indicative of possible changes in the cellular landscape. Although our study emphasises the potential of RPs in high-dimensional data monitoring, further examination is essential to understand the implications of these findings. If these variations are associated with changes in the health condition of patients (such as the onset of a disease), they could potentially serve as invaluable early warning signs for medical practitioners, thereby facilitating timely interventions.

VI. DISCUSSION

In this study, we examine the application and performance of a Multivariate Shewhart Control Chart using various RP methods on high-dimensional image data. As technological advancements have led to the rise of high-dimensional data, there is an increasing need to efficiently process, analyse, and control such data types in various fields. The relevance of this investigation lies in addressing this pressing need, specifically in the healthcare sector where high-dimensional image data are prevalent. This study aims to evaluate the impact of RP methods on the performance of the MSCC. Our extensive simulation studies vary RP methods and different parameters, including epsilon value, image size, and data volume.

The findings reveal several insights into how the MSCC performance is affected by shifts (or changes) in the parameters. As the magnitude of the shift increases, the ARL consistently decreases, implying that larger changes are more readily identified, which is intuitive because larger changes are easier to detect. Performance varies across different RP methods, with the PM method offering the most efficient results, particularly for smaller image sizes. Regarding the effect of reduced dimensions (d), increasing d leads to an increase in the ARL values, thereby slowing the detection of shifts. In addition, larger image sizes result in a slight increase in the ARL values, particularly for higher values of the shift and smaller values of d . This indicates that the monitoring process may be slower in detecting changes in larger images. These findings underscore the relevance and effectiveness of the MSCC approach using RP for high-dimensional imaging data. They also indicate the critical role of parameter selection in optimising the performance of these control charts.

Our study significantly contributes to the existing literature by addressing crucial research gaps in the application and performance assessment of MSCC using RP methods on high-dimensional image data. While previous studies have explored control charts [51] and image analysis [27] sep-

arately, there is a scarcity of research that systematically evaluates the impact of RP methods on the performance of MSCC in the context of high-dimensional data. Our study bridges this gap by providing a comprehensive analysis, bringing together control chart methodologies and advanced image processing techniques. Previous studies [17], [52], [53] emphasize the growing importance of such integrative approaches, underlining the need for sophisticated methodologies in the era of high-dimensional datasets.

The practical relevance of these findings is further underscored by their application in monitoring changes in the immune cell composition in images, particularly in healthcare applications such as the analysis of BAS and PMO cell images. By employing a combination of color, texture, and size features, including the distinctive size variable measured by the area of individual cells, our study offers essential insights into allergic reactions, autoimmune diseases, and acute promyelocytic leukaemia. The timely detection of shifts in these images can facilitate early diagnosis and intervention. Moreover, our observations highlight the influence of image size and reduced dimensions on the detection efficiency of RP methods, emphasizing the practical implications of the parameters studied in our simulation. Applying our findings to the analysis of cell images reveals valuable information, contributing to the overall effectiveness of our method in distinguishing between BAS and PMO cells in high-dimensional image data.

The potential applications of our study are far-reaching, particularly in domains where the effective monitoring and control of high-dimensional data are imperative. In healthcare, where high-dimensional image data are abundant, our methodology can empower practitioners with enhanced tools for early detection of anomalies in medical imaging, leading to more accurate diagnoses and personalized treatment plans. Moreover, the robustness of our approach extends to commerce [54] and manufacturing [55], where it can be applied to monitor and control complex production processes, ensuring product quality.

Although the study offers crucial insights into the application of MSCC to high-dimensional imaging data using RP methods, few limitations are noteworthy. The diversity and sample size of the images may constrain the generalisability of the findings. Although we explored various RP techniques, the expansive domain of dimensionality reduction may have been unaccounted for, potentially resulting in different outcomes. Computational constraints may also affect practical implementation, and our controlled environment does not capture the variability introduced by real-world scenarios such as equipment variations or environmental conditions. Finally, the complexity of real-world scenarios, particularly those with nonlinear shifts, may challenge the applicability of the findings. Despite these challenges, our method is a robust and promising approach that demonstrates significant potential for practical applications in monitoring and controlling high-dimensional image data.

While our study makes significant strides in understanding

the performance of MSCC with RP methods, there exist promising avenues for future research. Further research could explore additional dimensionality reduction methods, different types of image data, and other factors that might influence the control chart performance. Future studies might also consider more complex scenarios, such as multivariate shifts or nonlinear changes, which would further test the robustness and adaptability of MSCC with dimensionality reduction methods. Investigating the adaptability and scalability of our methodology to different types of high-dimensional data, such as hyperspectral imaging or volumetric data, presents an intriguing direction. Furthermore, exploring the potential integration of machine learning techniques to further enhance the predictive capabilities of our model represents a compelling area for future inquiry.

VII. CONCLUSION

This study highlights the practical application of dimensionality reduction methods with multivariate Shewhart control chart to control the quality of high-dimensional image data. We integrate MSCC with several random projection methods - Achlioptas, plus-minus one, Li and normal projections - to identify changes in immune cell composition in high-dimensional images. Through extensive simulation studies, we reveal nuanced insights into how the proposed scheme performance is influenced by varying RP methods and different parameters. Notably, larger shifts in images are more readily identified, with PM demonstrating optimal results among other RP methods. Uniquely, our method provides deeper insights into individual patient samples, allowing for a clearer understanding of group differences. This research serves as a stepping stone for future endeavors, encouraging the exploration of optimized parameter configurations and extending the application of MSCC with RP to diverse domains beyond healthcare.

REFERENCES

- [1] Edwin R Parra, Carmen Behrens, Jaime Rodriguez-Canales, Heather Lin, Barbara Mino, Jorge Blando, Jianjun Zhang, Don L Gibbons, John V Heymach, Boris Sepesi, et al. Image analysis-based assessment of pd-1 and tumor-associated immune cells density supports distinct intratumoral microenvironment groups in non-small cell lung carcinoma patients. *Clinical Cancer Research*, 22(24):6278–6289, 2016.
- [2] WT Chen, A Lu, K Craessaerts, B Pavie, C Sala Frigerio, N Corthout, X Qian, J Laláková, M Kühnemund, I Voytyuk, et al. Spatial transcriptomics and in situ sequencing to study alzheimer's disease. *cell*, 2020.
- [3] Tuba N Gide, Camelia Quek, Alexander M Menzies, Annie T Tasker, Ping Shang, Jeff Holst, Jason Madore, Su Yin Lim, Rebecca Velickovic, Matthew Wongchenko, et al. Distinct immune cell populations define response to anti-pd-1 monotherapy and anti-pd-1/anti-ctla-4 combined therapy. *Cancer cell*, 35(2):238–255, 2019.
- [4] Guray Akturk, Robert Sweeney, Romain Remark, Miriam Merad, and Sacha Gnjatich. Multiplexed immunohistochemical consecutive staining on single slide (micssss): multiplexed chromogenic ihc assay for high-dimensional tissue analysis. *Biomarkers for immunotherapy of cancer: methods and protocols*, pages 497–519, 2020.
- [5] Walter Andrew Shewhart. *Control of quality of manufactured product*. 1929.
- [6] Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & sons, 2019.
- [7] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250, 2001.
- [8] Laurens Van Der Maaten, Eric O Postma, H Jaap van den Herik, et al. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(66-71):13, 2009.
- [9] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003. Special Issue on PODS 2001.
- [10] Ping Li, Trevor J Hastie, and Kenneth W Church. Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 287–296, 2006.
- [11] Daniel W Apley and Jianjun Shi. The glrt for statistical process control of autocorrelated processes. *IIE transactions*, 31(12):1123–1134, 1999.
- [12] Kroopa Joshi, Marc Robert de Massy, Mazlina Ismail, James L Reading, Imran Uddin, Annemarie Woolston, Emine Hatipoglu, Theres Oakes, Rachel Rosenthal, Thomas Peacock, et al. Spatial heterogeneity of the t cell receptor repertoire reflects the mutational landscape in lung cancer. *Nature medicine*, 25(10):1549–1559, 2019.
- [13] William B Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, 1986.
- [14] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.
- [15] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003.
- [16] Santosh S Vempala. *The random projection method*, volume 65. American Mathematical Soc., 2005.
- [17] Ewa Skubalska-Rafajłowicz. Random projections and hotelling's t2 statistics for change detection in high-dimensional data streams. *International Journal of Applied Mathematics and Computer Science*, 23(2), 2013.
- [18] William B. Johnson. Extensions of lipschitz mappings into hilbert space. *Contemporary mathematics*, 26:189–206, 1984.
- [19] Dimitris Achlioptas and Frank Mcsherry. Fast computation of low-rank matrix approximations. *J. ACM*, 54(2):9–es, apr 2007.
- [20] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563, 2006.
- [21] Peter Buneman. *Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. ACM, 2001.
- [22] Xiaoli Z Fern and Carla E Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 186–193, 2003.
- [23] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *2006 47th annual IEEE symposium on foundations of computer science (FOCS'06)*, pages 143–152. IEEE, 2006.
- [24] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613, 1998.
- [25] Gregoire Pau, Florian Fuchs, Oleg Sklyar, Michael Boutros, and Wolfgang Huber. Ebimage—an r package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7):979–981, 2010.
- [26] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.
- [27] SP Tamizhselvi, P Revathy, I Ambika, E Manogar, G Renganayagi, and C Mahiba. Analysis of cancerous liver mri image using various segmentation methods. *International Journal of Intelligent Systems and Applications in Engineering*, 12(9s):524–528, 2024.
- [28] Jing-Lin Xiao, Jian-Sheng Fan, Yu-Fei Liu, Bao-Luo Li, and Jian-Guo Nie. Region of interest (roi) extraction and crack detection for uav-based bridge inspection using point cloud segmentation and 3d-to-2d projection. *Automation in Construction*, 158:105226, 2024.
- [29] Yanhui Guo, Ahmed I Shahin, and Harish Garg. An indeterminacy fusion of encoder-decoder network based on neutrosophic set for white blood cells segmentation. *Expert Systems with Applications*, page 123156, 2024.
- [30] Yinghao Yao, Jiaying Yang, Haojun Sun, Hengte Kong, Sheng Wang, Ke Xu, Wei Dai, Siyi Jiang, QingShi Bai, Shilai Xing, et al. Deepgraft: A novel semantic segmentation auxiliary roi-based deep learning framework for effective fundus tessellation classification. *Computers in Biology and Medicine*, 169:107881, 2024.

- [31] Getao Du, Xu Cao, Jimin Liang, Xueli Chen, and Yonghua Zhan. Medical image segmentation based on u-net: A review. *Journal of Imaging Science and Technology*, 2020.
- [32] Behnood Rasti, Danfeng Hong, Renlong Hang, Pedram Ghamisi, Xudong Kang, Jocelyn Chanussot, and Jon Atli Benediktsson. Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. *IEEE Geoscience and Remote Sensing Magazine*, 8(4):60–88, 2020.
- [33] Samuel Budd, Emma C Robinson, and Bernhard Kainz. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71:102062, 2021.
- [34] Jonas Windhager, Vito Riccardo Tomaso Zanotelli, Daniel Schulz, Lasse Meyer, Michelle Daniel, Bernd Bodenmiller, and Nils Eling. An end-to-end workflow for multiplexed image processing and analysis. *Nature Protocols*, 18(11):3565–3613, 2023.
- [35] Chuen-Sheng Cheng, Pei-Wen Chen, Yu-Chin Hsieh, and Yu-Tang Wu. Multivariate process control chart pattern classification using multi-channel deep convolutional neural networks. *Mathematics*, 11(15):3291, 2023.
- [36] Ridwan A Sanusi, Amitava Mukherjee, and Min Xie. A comparative study of some ewma schemes for simultaneous monitoring of mean and variance of a gaussian process. *Computers & Industrial Engineering*, 135:426–439, 2019.
- [37] Chenglong Li, Amitava Mukherjee, and Qin Su. A distribution-free phase i monitoring scheme for subgroup location and scale based on the multi-sample lepage statistic. *Computers & Industrial Engineering*, 129:259–273, 2019.
- [38] Nurudeen A Adegoke, Muhammad Riaz, Ridwan A Sanusi, Adam NH Smith, and Matthew DM Pawley. Ewma-type scheme for monitoring location parameter using auxiliary information. *Computers & Industrial Engineering*, 114:114–129, 2017.
- [39] IM Zwetsloot, LA Jones-Farmer, and WH Woodall. Monitoring univariate processes using control charts: Some practical issues and advice. *Quality Engineering*, pages 1–13, 2023.
- [40] Mahdijeh Erfanian, Bahram Sadeghpour Gildeh, and Mahmoud Reza Azarpazhooh. A new approach for monitoring healthcare performance using generalized additive profiles. *Journal of Statistical Computation and Simulation*, 91(1):167–179, 2021.
- [41] Ridwan A Sanusi, Muhammad Riaz, Nurudeen A Adegoke, and Min Xie. An ewma monitoring scheme with a single auxiliary variable for industrial processes. *Computers & Industrial Engineering*, 114:1–10, 2017.
- [42] Nurudeen A Adegoke, Adam NH Smith, Marti J Anderson, Saddam Akber Abbasi, and Matthew DM Pawley. Shrinkage estimates of covariance matrices to improve the performance of multivariate cumulative sum control charts. *Computers & Industrial Engineering*, 117:207–216, 2018.
- [43] Kelly D Stone, Calman Prussin, and Dean D Metcalfe. Ige, mast cells, basophils, and eosinophils. *Journal of Allergy and Clinical Immunology*, 125(2):S73–S80, 2010.
- [44] Hajime Karasuyama and Yoshinori Yamanishi. Basophils have emerged as a key player in immunity. *Current opinion in immunology*, 31:1–7, 2014.
- [45] H Jeffrey Lawrence and Corey Largman. Homeobox genes in normal hematopoiesis and leukemia. 1992.
- [46] Francesco Lo-Coco, Giuseppe Avvisati, Marco Vignetti, Christian Thiede, Sonia Maria Orlando, Simona Iacobelli, Felicetto Ferrara, Paola Fazi, Laura Cicconi, Eros Di Bona, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *New England Journal of Medicine*, 369(2):111–121, 2013.
- [47] Mark M Davis. A prescription for human immunology. *Immunity*, 29(6):835–838, 2008.
- [48] C. Matek, S. Schwarz, C. Marr, and K. Spiekermann. A single-cell morphological dataset of leukocytes from aml patients and non-malignant controls. <https://doi.org/10.7937/tcia.2019.36f5o9ld>, 2019. [Data set].
- [49] Christian Matek, Simone Schwarz, Karsten Spiekermann, and Carsten Marr. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nature Machine Intelligence*, 1(11):538–544, 2019.
- [50] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26:1045–1057, 2013.
- [51] Nurudeen A Adegoke, Abdaljbbar Dawod, Olatunde Adebayo Adeoti, Ridwan A Sanusi, and Saddam Akber Abbasi. Monitoring multivariate coefficient of variation for high-dimensional processes. *Quality and Reliability Engineering International*, 38(5):2606–2621, 2022.
- [52] George C Runger. Projections and the u 2 multivariate control chart. *Journal of Quality Technology*, 28(3):313–319, 1996.
- [53] Olha Bodnar and Wolfgang Schmid. Multivariate control charts based on a projection approach. *Allgemeines Statistisches Archiv*, 89:75–93, 2005.
- [54] Chunhua Wang, Dong Tang, Hairong Lin, Fei Yu, and Yichuang Sun. High-dimensional memristive neural network and its application in commercial data encryption communication. *Expert Systems with Applications*, 242:122513, 2024.
- [55] Chenang Liu, Zhenyu Kong, Suresh Babu, Chase Joslin, and James Ferguson. An integrated manifold learning approach for high-dimensional data feature extractions and its applications to online process monitoring of additive manufacturing. *IIE Transactions*, 53(11):1215–1230, 2021.



DR. RIDWAN A. SANUSI is serving as a Postdoctoral Fellow at the Interdisciplinary Research Center for Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals (KFUPM) Saudi Arabia. He completed his PhD at the City University of Hong Kong, Hong Kong, and obtained his M.Sc. in Applied Statistics from KFUPM, as well as a B.Sc. in Statistics from the University of Ibadan, Nigeria. His research interests include Pure & Applied Statistics, Statistical Quality Control, Machine Learning, Biostatistics, and High-dimensional Processes.



DR. JIMOH OLAWALE AJADI is serving as an Assistant Professor of Statistics in the Department of Mathematics, at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. He received his PhD from the City University of Hong Kong, and his M.Sc. in Applied Statistics from King Fahd University of Petroleum & Minerals. His research interests include statistical process monitoring and data science.



DR. SADDAM ABBASI is currently working as an Associate Professor of Statistics in the Department of Mathematics, Statistics and Physics, College of Arts and Science, Qatar University. He received his Ph.D degree in Statistic from The University of Auckland, New Zealand in 2013. Since then, he has served at prestigious institutions including 5 years experience at Qatar University. He is deeply involved with research in the field of quality control, data science and industrial engineering. His research work resulted in more than 100 ISI ranked journal articles with more than 2200 citations and an h-index of 27. He also acts as a reviewer for several international journals and conferences. Due to his outstanding research achievements, he is rated among the top 2% of world scientists in his research area.

TABLE 1. Performance of the RP-based MSCC, with reduced dimension (d) = 2

δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	198.39 (197.55)	202.96 (202.49)	196.91 (197.66)	199.07 (199.76)	199.55 (197.83)	198.74 (199.8)	198.35 (198.92)	203.17 (202.31)
0.025	163.47 (163.4)	162.89 (163.89)	155.97 (154.78)	154.66 (153.86)	164.28 (164.43)	159.94 (158.71)	158.16 (157.61)	161.63 (161.23)
0.05	135.19 (135.24)	131.12 (130.52)	124.86 (125.14)	121.5 (121.02)	136.56 (135.41)	127.07 (127.14)	128.31 (128.29)	128.1 (127.56)
0.075	112.97 (112.07)	103.41 (102.33)	100.12 (100.13)	96.02 (95.72)	114.58 (114.49)	104.6 (103.64)	104.51 (103.03)	104.01 (102.71)
0.1	93.08 (92.9)	83.88 (83.62)	81.21 (81.02)	76.91 (75.7)	97.04 (96.29)	86.12 (85.14)	85.57 (85.84)	84.75 (83.86)
0.2	45.13 (44.38)	37.64 (36.98)	36.49 (35.99)	33.68 (33.2)	48.96 (48.1)	40.93 (40.3)	40.57 (40.22)	39.75 (39.11)
0.4	14.16 (13.5)	11.66 (11.14)	11.44 (10.96)	10.79 (10.28)	15.93 (15.38)	12.94 (12.41)	13.07 (12.55)	12.44 (11.94)
0.6	6.61 (6.11)	5.61 (5.1)	5.5 (4.97)	5.21 (4.68)	7.19 (6.69)	6.13 (5.56)	6.08 (5.54)	5.84 (5.33)
0.8	3.97 (3.44)	3.54 (3.01)	3.49 (2.96)	3.35 (2.81)	4.25 (3.72)	3.75 (3.22)	3.72 (3.17)	3.65 (3.12)
1	2.86 (2.31)	2.59 (2.03)	2.58 (2.03)	2.5 (1.93)	2.98 (2.43)	2.72 (2.17)	2.7 (2.15)	2.65 (2.1)
2	1.81 (1.21)	1.78 (1.18)	1.8 (1.2)	1.75 (1.14)	1.84 (1.25)	1.81 (1.2)	1.84 (1.24)	1.85 (1.25)
3	1.82 (1.23)	1.79 (1.18)	1.79 (1.2)	1.76 (1.15)	1.84 (1.25)	1.82 (1.22)	1.84 (1.25)	1.85 (1.26)

* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.

TABLE 2. Performance of the RP-based MSCC, with reduced dimension (d) = 5

δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	199.25 (198.51)	198.18 (197.44)	204.41 (203.31)	200.1 (199.09)	201.75 (200.95)	200.88 (201.55)	200.49 (200.94)	201.34 (201.23)
0.025	168.3 (166.53)	161.77 (162.08)	168.98 (168.75)	161.74 (161.24)	173.21 (172.93)	167.55 (166.57)	167.9 (166.17)	163.77 (163.58)
0.05	144.93 (144.57)	136.05 (135.62)	137.85 (138.11)	131.06 (129.28)	146.16 (146.44)	138.43 (138.29)	139.13 (138.42)	135.83 (135.96)
0.075	122.14 (121.61)	112.05 (111.27)	114.94 (115.75)	108.16 (108.02)	124.83 (124.07)	116.24 (116.1)	116.67 (116.04)	112.38 (111.71)
0.1	104.99 (104.41)	93.5 (93.36)	95.75 (95.38)	88.51 (87.53)	108.2 (107.27)	98.21 (97.47)	98.04 (97.04)	94.37 (94.31)
0.2	55.55 (55.37)	45.98 (45.68)	47.09 (46.58)	42.91 (42.8)	61.19 (60.52)	51.59 (51.19)	51.98 (51.63)	48.94 (48.44)
0.4	18.41 (18)	14.45 (13.84)	14.65 (14.16)	13.37 (12.82)	21.35 (20.81)	17 (16.41)	17.1 (16.51)	15.92 (15.37)
0.6	7.92 (7.38)	6.42 (5.92)	6.47 (5.92)	6.03 (5.5)	9.18 (8.64)	7.49 (6.99)	7.48 (6.95)	7.03 (6.46)
0.8	4.35 (3.82)	3.67 (3.14)	3.68 (3.16)	3.46 (2.93)	4.92 (4.4)	4.11 (3.58)	4.14 (3.61)	3.94 (3.37)
1	2.87 (2.32)	2.5 (1.94)	2.51 (1.96)	2.41 (1.84)	3.14 (2.6)	2.73 (2.18)	2.75 (2.2)	2.65 (2.09)
2	1.26 (0.58)	1.21 (0.51)	1.22 (0.52)	1.2 (0.48)	1.28 (0.6)	1.23 (0.53)	1.23 (0.54)	1.22 (0.52)
3	1.19 (0.47)	1.15 (0.42)	1.15 (0.42)	1.14 (0.4)	1.2 (0.49)	1.17 (0.44)	1.17 (0.44)	1.16 (0.43)

* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.

TABLE 3. Performance of the RP-based MSCC, with reduced dimension (d) = 10

δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	199.28 (199.33)	196.85 (195.88)	200.91 (200.53)	201.31 (201.9)	197.77 (197.17)	199.83 (198.97)	202.33 (204.56)	202.06 (200.19)
0.025	171.62 (171.23)	164.07 (164.66)	169.8 (170.66)	168.33 (167.79)	170.16 (168.53)	169.65 (169)	170.78 (171.35)	170.97 (170.38)
0.05	148.16 (148.25)	137.9 (138.03)	142.8 (142.75)	141.03 (141.11)	148.84 (149.09)	142.82 (141.68)	145.43 (144.2)	143.48 (142.59)
0.075	127.27 (126.48)	117.66 (116.52)	121.26 (121.45)	117.86 (117.91)	128.30 (127.24)	120.03 (119.96)	122.66 (121.55)	120.19 (119.52)
0.1	110.87 (110.56)	99.35 (98.77)	103.05 (102.65)	97.82 (97.81)	111.31 (110.27)	103.44 (103.61)	104.57 (104.64)	101.24 (99.81)
0.2	62.51 (61.52)	52.82 (52.86)	54.19 (53.68)	50.33 (50.11)	66.37 (65.29)	58.26 (57.65)	58.37 (57.64)	55.11 (54.67)
0.4	21.88 (21.36)	17.38 (16.86)	17.74 (17.29)	16.36 (15.82)	24.98 (24.71)	20.44 (20.13)	20.65 (20.15)	19.34 (18.85)
0.6	9.29 (8.7)	7.39 (6.87)	7.48 (7.01)	7.00 (6.44)	10.94 (10.45)	8.81 (8.33)	8.92 (8.4)	8.45 (7.93)
0.8	4.87 (4.37)	3.99 (3.48)	4.05 (3.51)	3.80 (3.26)	5.64 (5.1)	4.67 (4.12)	4.71 (4.19)	4.48 (3.94)
1	3.02 (2.45)	2.58 (1.99)	2.6 (2.05)	2.48 (1.91)	3.40 (2.87)	2.91 (2.36)	2.92 (2.35)	2.83 (2.27)
2	1.18 (0.46)	1.14 (0.4)	1.14 (0.40)	1.13 (0.37)	1.22 (0.51)	1.17 (0.44)	1.17 (0.45)	1.16 (0.43)
3	1.03 (0.16)	1.02 (0.14)	1.02 (0.14)	1.02 (0.14)	1.03 (0.17)	1.02 (0.15)	1.02 (0.15)	1.02 (0.15)

* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.



DR. TAOFIK O. DAUDA is an Associate Professor of Biometrics at the Institute of Agricultural Research and Training, Obafemi Awolowo University, Nigeria. He earned his PhD in Statistics from the Federal University of Agriculture Abeokuta, Nigeria. He has served as a reviewer for various international journals and conferences. His research interests include spatial statistics, applied statistics, and information system development.



DR. NURUDEEN A. ADEGOKE is a Research Data Scientist at the Melanoma Institute Australia, affiliated with The University of Sydney. His expertise lies in a synergistic blend of machine learning, data mining, and analytics, all geared towards advancing our understanding of melanoma. Additionally, his work focuses on data science, biostatistics, and high-dimensional data analysis, demonstrating a dedication to using data science for medical advancements. His current research reflects a deep engagement with these interdisciplinary fields, aiming to drive innovations in melanoma treatment and prevention.

TABLE 4. Performance of the RP-based MSCC, with reduced dimension (d) = 15

δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	193.67 (193.46)	200.36 (200.85)	197.11 (195.62)	202.37 (202.07)	203.21 (204.42)	198.67 (198.96)	200.65 (199.6)	201.65 (202.27)
0.025	166.23 (165.88)	169.92 (169.92)	167.63 (167.32)	169.85 (169.62)	177.53 (179.12)	169.07 (168.27)	171.98 (172.18)	169.27 (167.77)
0.05	146.23 (145.61)	145.29 (145.93)	142.89 (141.82)	143.15 (142.8)	154.15 (152.8)	143.57 (143.44)	147.58 (145.62)	144.29 (144.5)
0.075	126.6 (125.9)	122.89 (122.41)	120.51 (120.43)	121.67 (121.32)	134.91 (133.96)	124.02 (123.31)	125.85 (125.34)	122.39 (121.68)
0.1	110.47 (109.13)	106.3 (105.52)	103.09 (102.76)	103.04 (103.43)	116.17 (115.1)	106.24 (106.45)	107.28 (107.52)	104.49 (103.76)
0.2	63.53 (63.31)	56.85 (56.64)	56.53 (56.3)	55.04 (54.52)	71.22 (70.95)	60.67 (60.17)	61.78 (60.99)	58.87 (57.92)
0.4	23.54 (23.02)	19.29 (18.82)	19.28 (18.71)	18.18 (17.78)	27.83 (27.16)	22.3 (21.86)	22.69 (22.23)	21.37 (20.94)
0.6	10 (9.5)	8.17 (7.68)	8.19 (7.63)	7.62 (7.11)	12.19 (11.65)	9.72 (9.28)	9.76 (9.24)	9.29 (8.77)
0.8	5.12 (4.63)	4.25 (3.76)	4.23 (3.72)	4.04 (3.52)	6.28 (5.7)	5.04 (4.5)	5.09 (4.58)	4.84 (4.34)
1	3.14 (2.57)	2.69 (2.12)	2.68 (2.11)	2.58 (2)	3.7 (3.17)	3.09 (2.55)	3.12 (2.57)	2.96 (2.42)
2	1.15 (0.41)	1.12 (0.36)	1.11 (0.35)	1.11 (0.34)	1.19 (0.47)	1.15 (0.41)	1.14 (0.4)	1.14 (0.4)
3	1.01 (0.12)	1.01 (0.1)	1.01 (0.1)	1.01 (0.1)	1.02 (0.14)	1.01 (0.12)	1.01 (0.11)	1.01 (0.11)

* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.

TABLE 5. Performance of the RP-based MSCC, with reduced dimension (d) = 20

δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	201.27 (200.87)	199.09 (199.18)	205.4 (206.04)	201.89 (200.06)	198.13 (196.91)	203.71 (205.23)	201.52 (201.46)	198.57 (196.11)
0.025	175.82 (174.96)	170.46 (170.09)	175.82 (173.65)	171.48 (170.06)	171.7 (172.82)	173.54 (172.9)	170.92 (168.56)	171.59 (170.16)
0.05	153.74 (153.56)	144.79 (144.91)	149.07 (148.5)	144.3 (143.73)	148.23 (148.09)	148.01 (147.79)	147.74 (147.24)	144.13 (143.7)
0.075	132.5 (131.25)	124.96 (124.64)	127.09 (126.56)	123.16 (122.47)	130.18 (129.15)	129.07 (127.53)	128.58 (126.74)	124.07 (124.23)
0.1	116.62 (117.25)	107.21 (106.45)	110.48 (109.2)	104.28 (104.07)	115.36 (115.11)	109.83 (108.65)	109.85 (109.54)	106.14 (106.22)
0.2	68.51 (67.64)	58.96 (57.88)	60.92 (60.21)	57.1 (56.55)	70.31 (69.35)	63.61 (62.87)	63.85 (63.23)	60.43 (59.94)
0.4	25.41 (24.8)	20.72 (20.28)	21.03 (20.61)	19.32 (18.86)	28.27 (27.72)	23.97 (23.31)	23.73 (23.12)	22.51 (22.06)
0.6	10.9 (10.33)	8.74 (8.24)	8.83 (8.33)	8.15 (7.66)	12.71 (12.19)	10.4 (9.85)	10.53 (10.04)	9.78 (9.21)
0.8	5.49 (4.99)	4.49 (3.95)	4.55 (4.01)	4.22 (3.67)	6.55 (6.01)	5.44 (4.93)	5.38 (4.84)	5.13 (4.58)
1	3.25 (2.71)	2.76 (2.2)	2.77 (2.22)	2.63 (2.08)	3.84 (3.3)	3.26 (2.69)	3.24 (2.68)	3.08 (2.56)
2	1.14 (0.4)	1.1 (0.33)	1.1 (0.34)	1.09 (0.32)	1.18 (0.46)	1.14 (0.39)	1.14 (0.4)	1.13 (0.38)
3	1.01 (0.1)	1.01 (0.08)	1.01 (0.08)	1.01 (0.07)	1.01 (0.11)	1.01 (0.1)	1.01 (0.09)	1.01 (0.09)

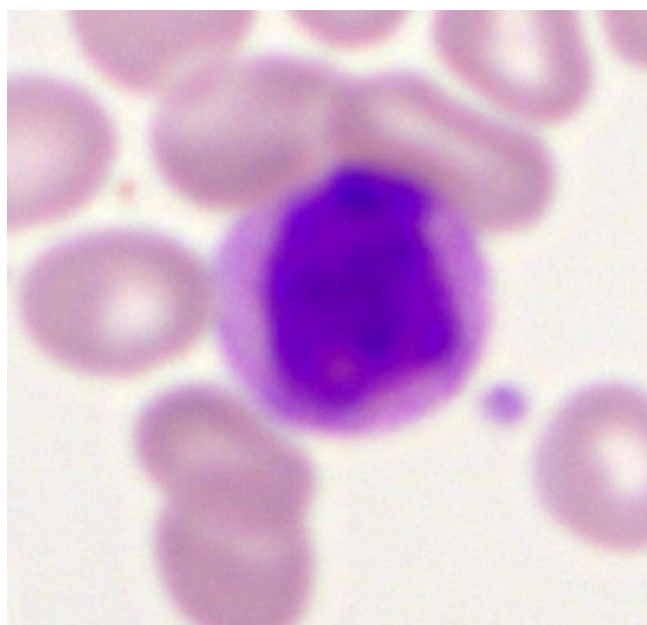
* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.

TABLE 6. Performance of the RP-based MSCC, with reduced dimension (d) = 30

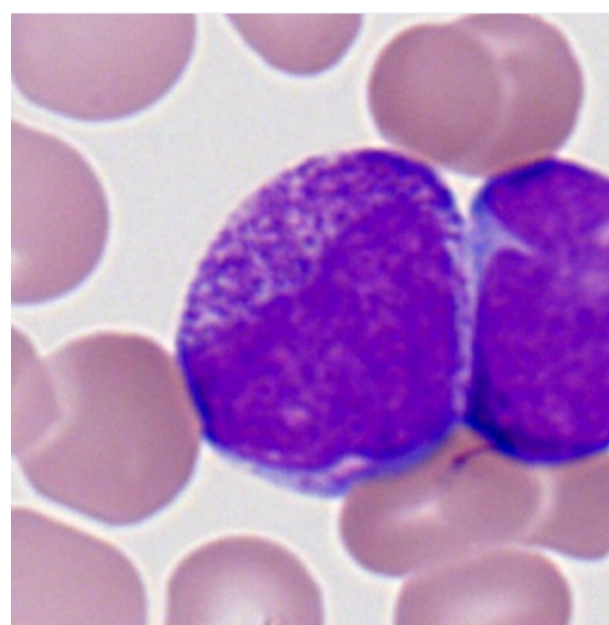
δ	image size 100 x 100				image size 142 x 142			
	Li	AP	NP	PM	Li	AP	NP	PM
	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)	ARL (SDRL)
0	202.86 (202.03)	197.36 (197.06)	206.2 (206.15)	203.78 (202.67)	200.36 (199.02)	195.75 (193.7)	201.59 (201.24)	200.01 (199.87)
0.025	175.63 (175.62)	169.33 (170.46)	176.49 (174.68)	174.35 (174.01)	175.81 (176.52)	170.92 (170.08)	175.33 (175.21)	172.42 (173.55)
0.05	154.83 (155.52)	145.83 (144.29)	152 (151.16)	149.48 (148.57)	153.92 (155.07)	145.5 (146.42)	151.09 (150.43)	148.74 (147.97)
0.075	135.56 (135.33)	126.45 (125.36)	130.29 (130.65)	127.43 (127.46)	135.21 (134.66)	126.51 (126.12)	130.66 (130.46)	128.51 (127.55)
0.1	118.15 (117.45)	108.42 (107.8)	113.68 (113.17)	109.14 (108.55)	119.09 (118.88)	109.31 (109.01)	113.03 (112.36)	111.78 (111.02)
0.2	71.26 (70.43)	62.09 (61.21)	64.03 (63.28)	60.87 (60.33)	72.98 (72.65)	64.2 (63.47)	66 (65.6)	63.43 (63.29)
0.4	26.9 (26.39)	21.99 (21.54)	22.86 (22.56)	21.21 (20.67)	30.33 (29.62)	25.02 (24.46)	25.88 (25.36)	24.53 (23.97)
0.6	11.84 (11.28)	9.46 (8.96)	9.72 (9.29)	8.89 (8.43)	13.83 (13.29)	11.13 (10.7)	11.48 (10.94)	10.9 (10.38)
0.8	5.93 (5.4)	4.8 (4.33)	4.91 (4.36)	4.55 (4.02)	7.06 (6.52)	5.74 (5.2)	5.88 (5.36)	5.6 (5.06)
1	3.46 (2.93)	2.88 (2.32)	2.93 (2.39)	2.76 (2.2)	4.13 (3.57)	3.43 (2.87)	3.5 (2.96)	3.34 (2.77)
2	1.13 (0.38)	1.09 (0.31)	1.10 (0.32)	1.08 (0.3)	1.18 (0.46)	1.13 (0.39)	1.13 (0.39)	1.12 (0.36)
3	1.00 (0.07)	1.00 (0.06)	1.00 (0.06)	1.00 (0.05)	1.01 (0.09)	1.01 (0.08)	1.01 (0.07)	1.00 (0.07)

* δ - shift, ARL - Average Run Length, SDRL - Standard deviation run length, Li projection, AP - Achlioptas projection, NP - Normal projection, PM - Plus-Minus One projection.

...



Sample Basophil (BAS) cell image.



Sample Promyelocyte (PMO) cell image.

FIGURE 1. Illustrative examples of the distinct cell images employed in our research. Originally captured as 400×400 pixel images with three channels, these samples were processed by averaging across all channels, resulting in single-channel images. These provide insights into the morphological characteristics of BAS and PMO cells, which form the foundation of our analysis.

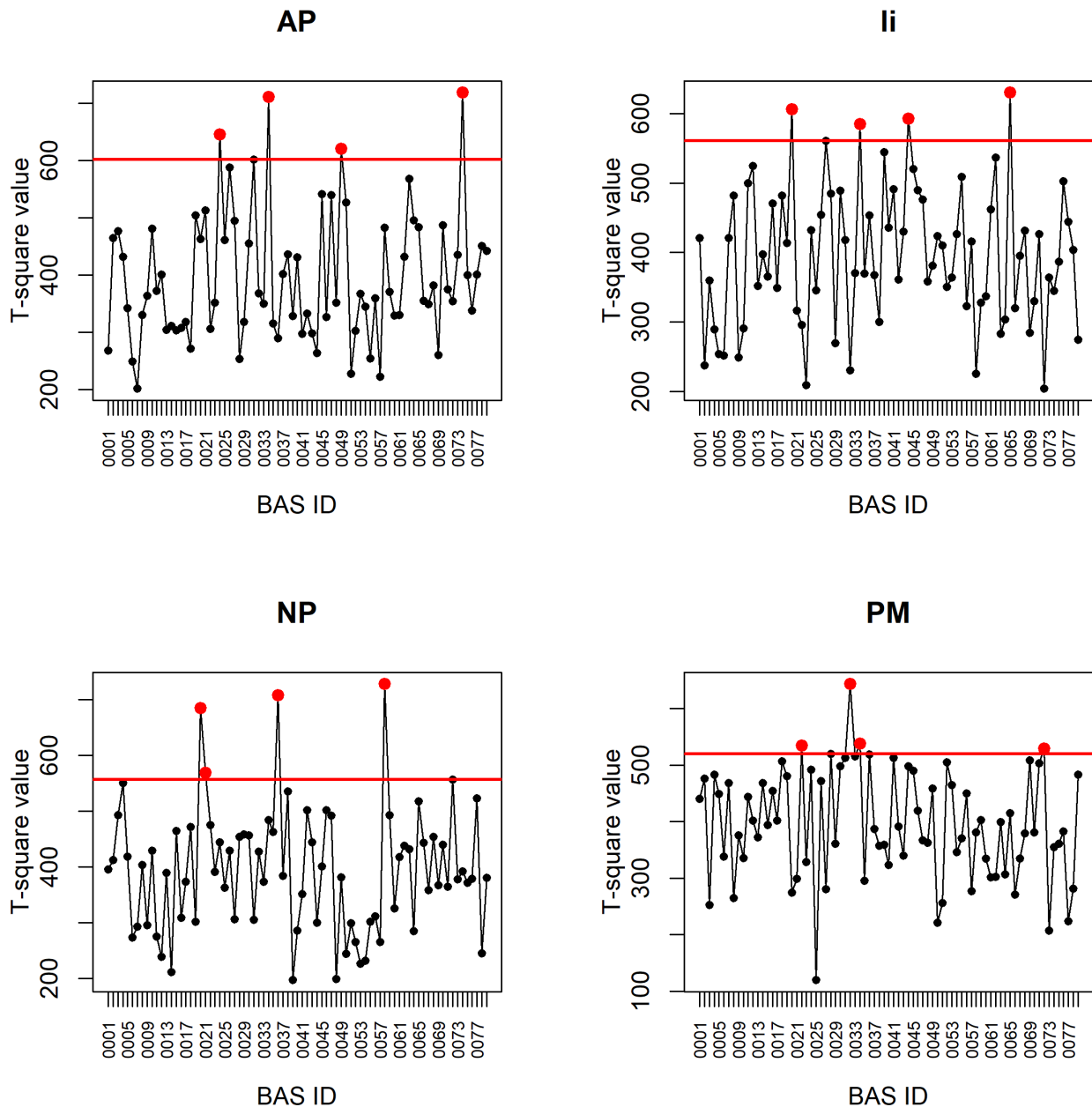


FIGURE 2. Initial RP-based MSCC Results from Phase I (BAS Images): All four methods (Li, AP, NP, and PM) consistently identifying four outlier samples beyond the upper control limits (UCL).

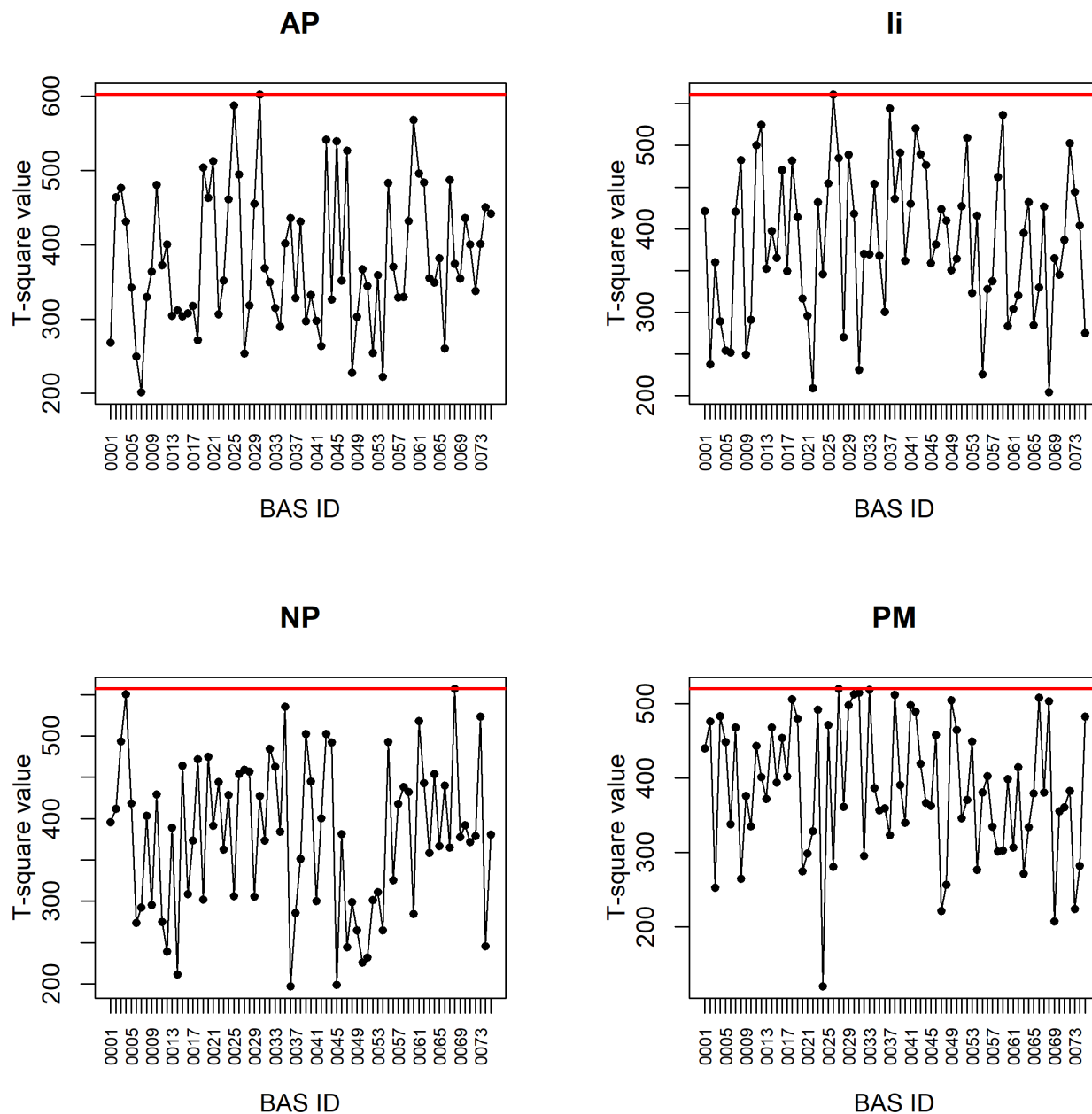


FIGURE 3. Refined Phase I RP-based MSCC Results (BAS Images): Post-outlier removal, illustrating a stabilization with no samples exceeding the UCL across all projection methods.

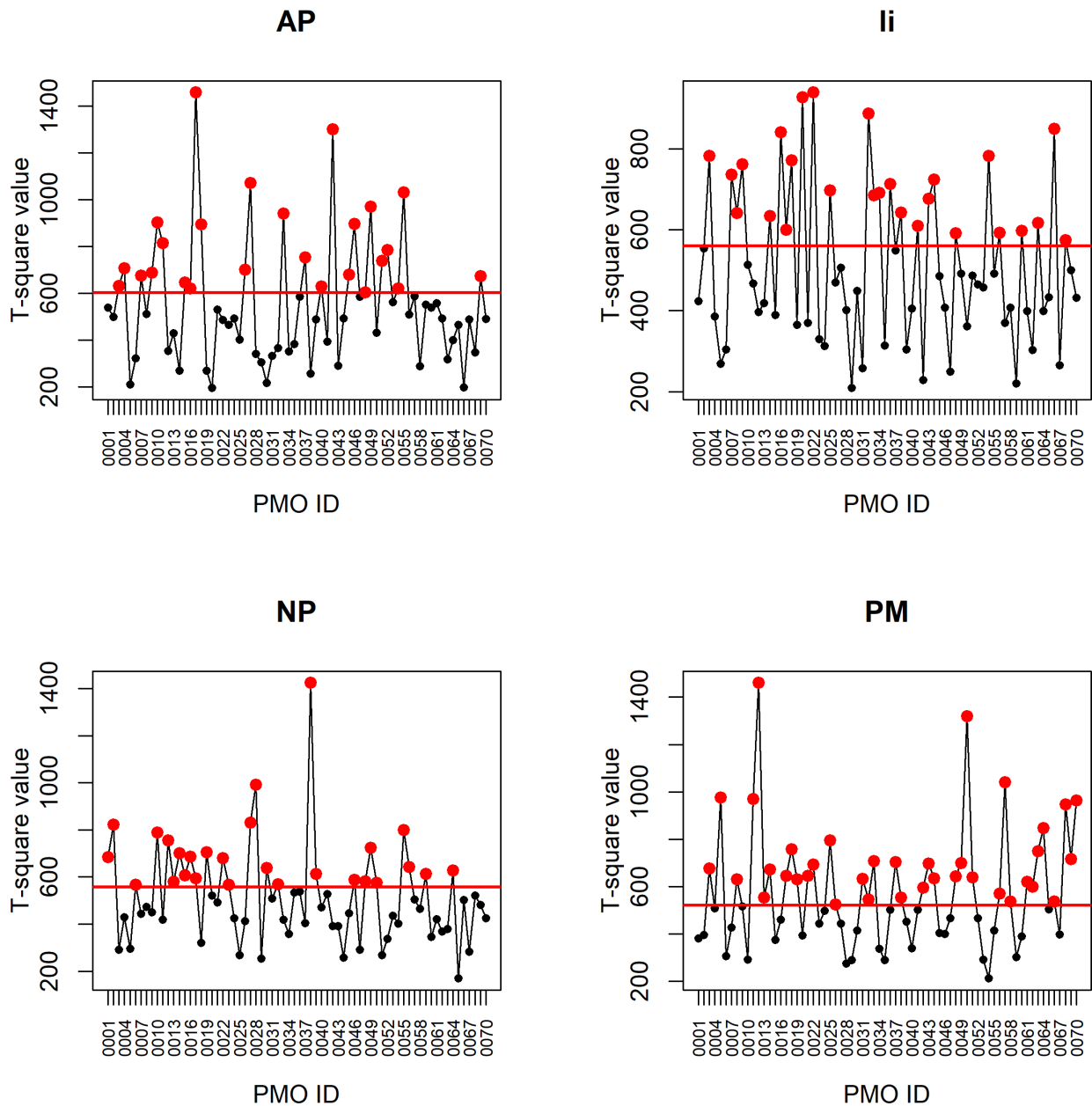


FIGURE 4. RP-based MSCC Results from Phase II (PMO Images): Displaying the number of samples beyond the UCL for each method. Specifically, PM identified 37 deviations, AP captured 25, NP flagged 27, and Li detected 26. PM showcases the highest deviation, underscoring its potential efficacy in detecting shifts in immune composition.