Superpixel Segmentation-Enabled Transmission Electron Microscopy Images for Rapid and Accurate Detection of Coronavirus

Bakr Ahmed Taha[a], Ros Maria Mat Yeh[a], Nurfarhana Mohd Sapiee[a], Yousif Al Mashhadany[b], Adawiya J. Haider[c], Mohd Hadri Hafiz Mokhtar[a] & Norhana Arsad[a*]

[a]*Photonic Technology Laboratory, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, UKM Bangi 43600, Malaysia*

[b]*Department of Electrical Engineering, College of Engineering, University of Anbar, Anbar,00964, Iraq;*

[c]*Applied Sciences Department/Laser Science and Technology Branch, University of Technology, Iraq.*

[*]*Corresponding author: noa@ukm.edu.my*

ABSTRACT

Worldwide, SARS-CoV-2 has been responsible for millions of fatalities and extensive disability. Hence, to stop the spread of novel viruses like SARS-CoV-2, Omicron, and other worrying types, rapid and accurate diagnostic techniques are needed to identify symptomatic and asymptomatic carriers as soon as feasible. Early recognition and diagnosis are essential to effective epidemic management. However, different viral strains' shapes and spatial characteristics are similar, complicating image classification, especially in medical virology. This study uses a super-pixels segmentation technique based on transmission electron microscopy (TEM) images to differentiate SARS-CoV-2 from SARS-CoV. This paper aims to develop a method that enables virologists to detect and diagnose viral infections more accurately. In results, SARS-CoV-2 had a median area of 25,145.54 pixels and SARS-CoV of 38,591.35 pixels. The model can help to better understand how viruses develop, spread, diagnose and contain outbreaks. Furthermore, an exceptionally low root mean square error (RMSE) of 0.0275 between the segmentation of the viral area between humans and machines is obtained. Indeed, this low error rate indicates the accuracy of this automated measurement technique. Finally, the developed superpixel segmentation technique provides quick and reliable identification of coronaviruses, promising to significantly contribute to medical virology and help manage epidemics by simplifying prompt viral diagnosis.

Keywords: SARS-CoV-2, SARS-COV; artificial intelligence; superpixel segmentation; Transmission Electron Microscopy

INTRODUCTION

Many people have lost their jobs and businesses have had to cut operations due to the ongoing global Coronavirus Disease 2019 (COVID-19) outbreak. According to data, 6,911,896 million people have died from contracting the virus and more than 694 million have been infected. The rapid spread of the virus negatively impacts public spaces, markets, schools, and airports, despite efforts to stop its spread. The rapid and widespread reach of the virus has highlighted the need for a more efficient environmental detection system to control its further spread and improve the health care of sufferers (Ai et al. 2020; Engineering 2021; Rahman et al. 2023; Bakr Ahmed Taha et al. 2020; Bakr Ahmed Taha, Al Mashhadany, et al. 2021; Bakr Ahmed Taha, Al-Jubouri, et al. 2022; Bakr Ahmed Taha, Ali, et al. 2021; Bakr Ahmed Taha, Mehde, et al. 2022)

Artificial intelligence (AI) encompasses methods that enable machines to imitate or exceed human intelligence, particularly in cognitive functions. The main branches of AI include machine learning, computer vision, and natural language processing (Joonas 2020; Mamat et al. 2023; Manea et al. 2023; Robertson et al. 2018; Bakr Ahmed

Taha, Al Mashhadany, Al-Jubouri, Haider, et al. 2023; Bakr Ahmed Taha, Al Mashhadany, Al-Jubouri, Rashid, et al. 2023a).

The machine learning techniques have frequently been employed to segment microorganisms present in the environment. Machine learning can be categorized into traditional methods and artificial neural networks. Previously, algorithms such as Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Randon Forest (RF), and others have been utilized to detection of viruses.

AI can potentially expedite even the most time-consuming tasks, making it particularly valuable in microorganism image analysis (Abd Alkarim et al. 2023; Li et al. 2015; Quality 2023; Rashid et al. 2023).

Researchers recommend using various technologies for virus identification, such as intelligent computer systems that implement deep learning algorithms to classify and organize microscopic images related to medical and food components (Fang et al. 2020; Kang et al. 2023; Kohler & Farr 1966; Rivenson et al. 2019; Schwartz 2020). Radiographs and Computed Tomography (CT) scans have been lauded as useful diagnostic tools in recent research on the diagnosis of respiratory disorders such as COVID-19. These methods are particularly helpful in determining the severity of COVID-19 infections, keeping tabs on the most seriously ill patients, and forecasting the course of the disease.

However, in such emergencies, relying solely on conventional manual diagnostic procedures is often impractical. One such answer is provided by computer-aided detection systems that use deep learning algorithms to make more accurate diagnoses in less time (Maghdid et al. 2021; Sahiner et al. 2019; Shi et al. 2021; Bakr A. Taha 2021).

Transmission Electron Microscopy (TEM) imagery is essential for studying significant viral replication and discovering new infections to create efficient plans for illness prevention, accurate diagnosis, control of viral outbreaks, and understanding the biology of pathogens and the causes of viral diseases (Richert-Pöggeler et al. 2019). Researchers have been trying to decipher viral architecture since the late 19th century, when it became clear that viruses played a significant role in disease transmission. The TEM image resolution is well-suited for nanoscale research. Therefore, it has the potential to give direct images of viruses for diagnostic and investigative reasons (Athirah et al. 2023; Mettenleiter 2017).

Although studies show that Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is 80% similar to Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), the two envelope viruses differ in the spike area's size, length, and density. These structural differences significantly affect the way these viruses are identified and diagnosed (Z. Zhu et al. 2020). In addition, coronavirus structures and morphologies observed in Vero cell cultures were determined using conventional thin-section TEMs with particle sizes of a median of 100 nm (Laue et al. 2021). Zhao et al. used TEM images to detect any differences in cell shape caused by SARS-CoV-2 to understand how the virus causes disease (Zhao et al. 2020).

A scarcity of literature is present on the topic of automated virus recognition in TEM images, with a disproportionate number of contributions coming from a limited group of researchers (Gustaf Kylberg et al. 2011; Matuszewski & Sintorn 2018; Roingeard et al. 2019). The researchers presented an automatic segmentation method to detect different types of virus particles in TEM images. The process consists of analyzing the local neighborhood of all pixels in the image, a broader object distinction, and a border correction phase for extended objects (G. Kylberg et al. 2012).

Image labels have been studied at the pixel level; however, deep learning methods have shown poor performance in the separation of visual objects. In particular, deep learning methods do not explicitly consider the possibility that neighboring pixels belong to the same category of objects in images (De Geus et al. 2019; Jha et al. 2020; Siam et al. 2018). Tracking Coronavirus particles in TEM images and calculating missed detection probabilities with deep recurrent neural networks Multiple recognize were combined to generate assignment probabilities. Moreover, the network calculated the likelihood that tracks would exist if created and destroyed (Spilger et al. 2020; Bakr Ahmed Taha, Al Mashhadany, Al-Jubouri, Rashid, et al. 2023b; Bakr Ahmed Taha, Al-Jubouri, et al. 2023; Bakr Ahmed Taha, Mashhadany, et al. 2022).

MORPHOLOGY MEASUREMENT

A morphology measurement method assigns numerical values to the geometric properties of an object in image processing and computer vision. The process involves assessing the spatial distribution of things and their relative sizes and shapes (Al-Kinani et al. 2020),(Yousif et al. 2012). It can identify these patterns from photos, and data can be extracted using a categorized approach. Alternatively, it can count objects, gauge their sizes, and identify their shapes from images. As part of this study, the threshold image's brightness and contrast were altered to reveal the details of the envelope and spike proteins. SARS-CoV-2 has a diameter of between 60 and 140 nm, making it spherical and giving it the appearance of a solar corona when viewed under an TEM (Ogando et al. 2020; Ou et

al. 2020; N. Zhu et al. 2020). In many studies, particle sizes use the geometric selection tools in (Image J software) such as freehand and segmented lines, to determine the outer viral envelope and length of spikes (Abràmoff et al. 2004; Laue et al. 2021; Yao et al. 2020).

With area selections, it becomes feasible to calculate the (max/min) diameters for virus features such as roundness, shape descriptors, diameter, area, and circularity. Subsequently, a stepwise extension in nanometers of the geometrical selection employed to identify the maximal diameter of each virus particle is employed to identify spikes associated with that particle. The accuracy of the technique was then verified using the adjusted scale bar tool to verify the spike length and diameters of various particles from various images, which was necessary to establish the average envelope diameters for both viruses. Moreover, the diameters of the virus particles were measured using a cross form since fixation artifacts in clinical specimens caused their noncircular shapes. Images of virus particles derived from cell cultures successfully mitigated the morphological variability resulting from fixation artifacts (Aggarwal et al. 2012).

Virus particle areas were determined using a free-form selection tool in the software. It is possible to decide on the location, line length, angle, and point values using this method based on the type of selection made (T. Ferreira 2012). Perform the analysis of particle areas by selecting a specific region and scanning the section until the outline of a virus shape is detected and repeat the process until the boundary of the selection allows for the extraction of the characteristics of the virus particles in a precise manner.

Figure 1 shows Coronavirus particle morphology based on TEM images. In Figure 1(A), the SARS-CoV virus can be seen, while in Figure 1(B), the SARS-CoV-2 virus can be seen with its scale bar adjusted to 100 nanometers. A consistent scale bar allows accurate measurement of virus particle sizes and interpretation of the results based on a precise comparison between the two images.
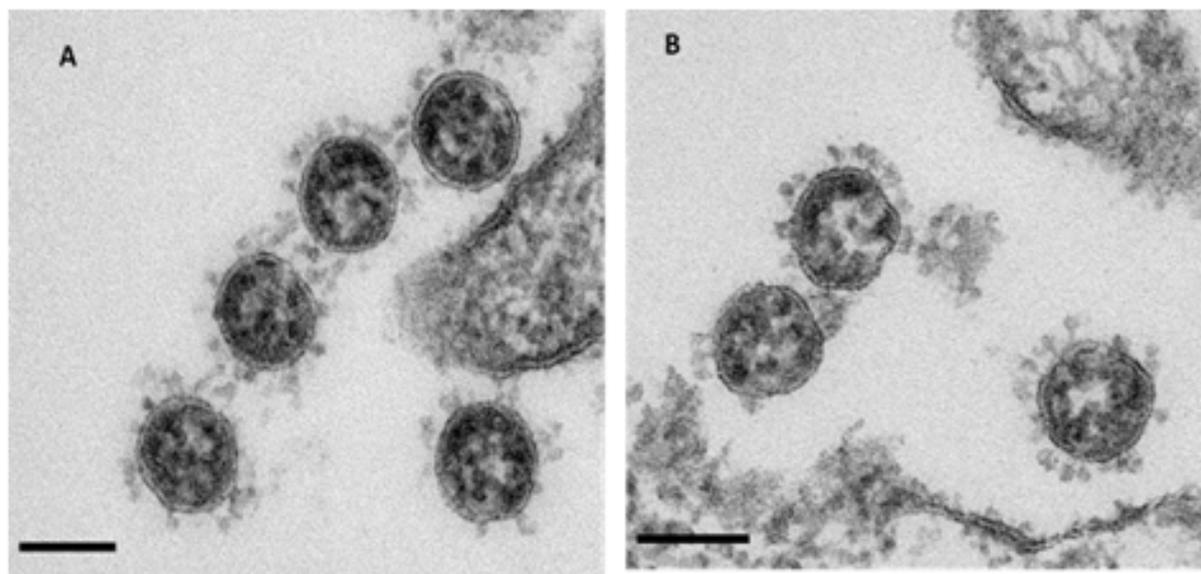

FIGURE 1. Illustrates a measurement of virus particle size in TEM images of Vero cells infected: A. SARS-CoV and B. SARS-CoV-2 at a Scale bar of 100 nm.

MATERIALS AND METHODS

Super-pixels are fundamental to identifying and acquiring visual features representing objects in future research activities that must meet the following characteristics: First, the process must be fast enough to be used in near-real-time scenarios. Second, clusters should display a low level of intra-cluster variance while displaying a high level of inter-cluster variance. Third, Groups should be positioned consistently within the imaging area. Finally, segmented

super-pixel boundaries are defined accurately by matching appropriately sized borders and discrete pixels. Figure 2 demonstrates the automated extraction process of area features for SARS-CoV-2 and SARS-CoV using the superpixel algorithm techniques. Further, emphasis is placed on preserving boundary information and obtaining localized data, which is essential for robust feature extraction and quantitative analysis in this work.
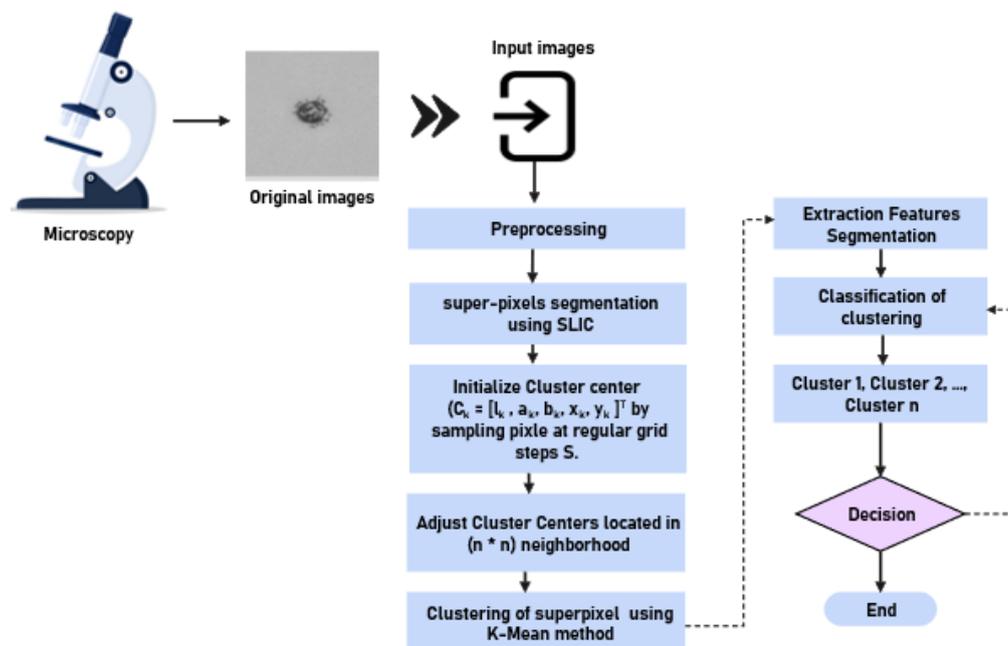

FIGURE 2. Flow chart of estimating the density of spike proteins and areas of SARS-CoV-2 using the superpixel segmentation technique

DATASET COLLECTION

Dataset files accessible on the Zenodo platform are classified into two groups, including 519 TEM images of the SARS-CoV 2 virus isolated from patients in Italy at INMI (Colavita et al. 2020), and 248 of the SARS-CoV virus isolated from patients in Frankfurt, Germany (Thiel et al. 2003). The image specification was a resolution (1376 x 1032) pixels, 0.64 nm, and a 16-bit TIF file. Irfan View software is necessary to view these images, but if you use ImageJ or Fiji software, Bio-format importers automatically apply appropriate measurements to files.

PRE-PROCESSING

Image enhancement plays a crucial role in the field of image analysis. Contrast enhancement is one of its techniques, as it improves image quality, tailoring it to specific needs. The primary objective of image enhancement revolves around rendering digital images more compatible with visual systems, achieved through actions such as smoothing or amplifying image details. Furthermore, this constitutes a notable focus within image analysis research investigations that have underscored the importance of histogram equalization (HE) as a fundamental and user-friendly method to increase contrast and overall image quality (Khan et.al 2015; Vidyasaraswathi 2020). In scientific imaging, spatial correlation is more significant than signal strength in discerning genetic materials. A low signal-to-noise ratio (SNR) often complicates the visual identification process. Although histogram equalization can sometimes yield misleading results, it is highly advantageous for scientific images, such as thermal, satellite, or X-ray images. Additionally, the histogram's intersection match value improves when one pixel matches the image color within the object, and there are fewer such colored pixels than within the image. As with normalizing histograms, normalizing them involves converting discrete intensity distributions into discrete probability distributions. Enhancing contrast in image histograms involves morphological mathematical operations, such as histogram normalization and equalization, which contribute to edge enhancement while mitigating noise.

SIMPLE LINEAR ITERATIVE CLUSTERING

Superpixel techniques are widely used in computer vision applications. They use set criteria to process all pixels randomly, resulting in unnecessary superpixel boundaries and regularity (Yuan et al. 2021). SLIC is better than the pixel-level approach. Due to this, the SLIC algorithm has been recently used to segment medical images (Cong et al. 2014). Image segmentation can be considered a set of boundaries or segments that span an entire image. The colour, brightness, and texture can be the same for the majority of the segmented pixels but differ dramatically between adjacent segments. Segment borders and edges tend to be closely connected due to the considerable brightness shifts at the section boundaries. Therefore, a different segmentation technique has been developed based on edge extraction methods using k-means clustering and SLIC segmentation to group nearby pixels. Initially, the desired number of superpixels, K, is chosen, with the aim of creating approximately equally sized superpixels. In an image with N pixels, the approximate size of each superpixel is $N = K$ pixels. To create superpixels of approximately equal size, a superpixel centre is chosen at

every grid interval as in the equation $S = \sqrt{N/k}$. In the k-means algorithm, clusters are divided based on the proximity of data points to the cluster centroids (C_k). Data points are assigned to the cluster whose centroid is closest to them in terms of Euclidean distance. This assignment is done iteratively until convergence, with data points potentially shifting between clusters to find the most appropriate grouping. At the beginning of the process, K superpixel cluster centres are selected as in the equation $C_k = \begin{bmatrix} l_k, & a_k, b_k, x_k, y_k \end{bmatrix}^T$ with $k = [1, K]$ at regular grid intervals SS. Because the spatial span of each superpixel is approximately S^2, it is assumed that pixels associated with a particular cluster centre are located on the xy plane within a $2S \times 2S$ radius around the superpixel centre. In contrast, there is the radius around the superpixel centre. This represents the search region for pixels that are closest to each cluster centre, and it is defined in the CIELAB colour space as well as the xy pixel coordinates. The SLIC algorithm uses the CIELAB colour space because it is perceptually homogeneous for tiny colour differences. The distance d_s between two pixels I and j in SLIC is a composite of two distances, dc and ds, which express colour and spatial closeness, respectively, as shown below in Eq. 1, Eq. 2, and Eq. 3 (Achanta et al. 2010)

$$d_c = \sqrt{\left(l_i - l_j\right)^2 + \left(a_i - a_j\right)^2 + \left(b_i - b_j\right)^2} \qquad (1)$$

$$d_s = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \qquad (2)$$

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \qquad (3)$$

D is denote the sum of the distances in both the lab and xy planes, which are then normalized by the grid spacing (S). The density of a superpixel (d_s) can be controlled by incorporating a variable m with the set higher emphasis on spatial proximity, thereby influencing the density of the cluster and giving added importance to the pixel pitch metric. d'_s denotes the indication of edges that exist between two pixels, suggesting the likelihood of an object boundary lying between these locations. To initiate the process of generating super-pixels, an edge detection algorithm assigns a value pi for each pixel I, indicating the probability of it being situated on a boundary. Consequently, the distance between two pixels is determined as the highest edge probability among all pixels

along the line connecting these pixels. Equations 4, 5, and 6 provide the revised distance calculation:

$$d'_c = \sqrt{\left(c_i - c_j\right)^2 + \left(y_i - y_j\right)^2 + \left(r_i - r_j\right)^2} \qquad (4)$$

$$d_s = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \qquad (5)$$

$$D' = \sqrt{(d'_c/m)^2 + (d_s/S)^2 + \alpha \times (d_e)^2} \qquad (6)$$

Where d'_c is the color distance between two pixels in an image with their color values red (r), green (g), and blue (b) color values. Squared differences between color values in each channel produce a scalar value, and the square root of this value is taken to obtain the color distance. d_s is the spatial distance between two pixels (i, j) in an image based on their x, y coordinate values in the image. The complete distance D' between two pixelswithin an image is established considering their color difference d'_c, spatial distance d_s, and an extra component $\alpha \times (d_e)^2$ which signifies the edge information's strength between the pixels. Scaling spatial distance is performed using a factor S, and normalizing color differences is accomplished utilizing a factor m.

Holistic distance is determined by taking the square root of the sum of squared color difference, spatial separation, and edge component. To ensure the creation of coherent super-pixels, all super-pixels within the image need to be interconnected to form edges. In the context of generating super-pixels, the decision to merge two neighboring super-pixels is guided by a similarity metric termed the chi-square (X^2) histogram distance in Equation 7 represents the disparity between two histograms denoted P and Q.

$$X^2(P, Q) = \frac{1}{2} \sum_k \frac{(P_k + Q_k)^2}{P_k - Q_k} \tag{7}$$
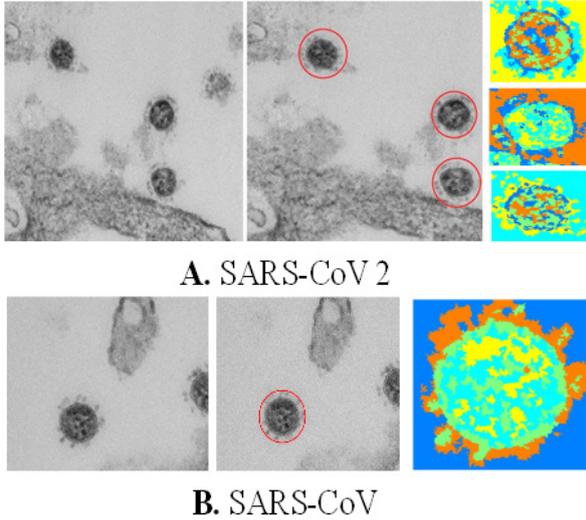

A. SARS-CoV 2

B. SARS-CoV

FIGURE 3. Identification of the superpixel segmentation based on the SLIC algorithm: A. SARS-COV 2 images, B. SARS-COV images

On the contrary, a superpixel-based segmentation approach involves partitioning images into regions corresponding to distinct objects or semantic sections within the images. The objective is to establish super-pixels that align with object boundaries and significant image features. These super-pixels can be generated through algorithms that optimize energy functions, including techniques like normalized cuts and SLIC, utilizing graphical representations of images. Figure 3 demonstrates the identification and detection of coronaviruses based on super-pixel segmentation from TEM images using the SLIC algorithm within MATLAB 2020.

K-MEAN CLUSTERING

In machine learning, the k-means clustering algorithm is often used for data clustering. In the selected clusters, the data automatically aggregated into numerous categories, with high similarities within each set and low similarities between them. Cluster-based super-pixels emerge by amalgamating similar pixels via clustering algorithms like k-means and mean shift. The notion entails partitioning the image into regions marked by similar colors, textures, or other pixel properties. However, clustering algorithms must align the borders between adjacent super-pixels to preserve the boundaries and accurately capture fine details. Post-processing techniques, such as graph-based optimization or edge-based refinement, can achieve this alignment process (Kumar et al. 2016). In the past, researchers have used a k-means-based multi-objective text clustering strategy. Based on the results, the proposed method performed better than the alternatives in text clustering, measured by well-known metrics such as precision and the F measure (Abualigah et al. 2016). The method was thoroughly tested on large datasets and consistently demonstrated its superior performance. The versatility and efficiency of the algorithm make it suitable for a wide range of text clustering tasks (Z. Chen et al. 2020; Luo 2016). Cluster K organizes the N-D data sets and forms meaningful clusters and the following pseudo-code form for the super-pixels algorithm process:

1. Initial specifying cluster centers, denoted as $m_1 \ldots \ldots . m_k \in IR^n$.

2. For each $x_i \in IR^n$ assign it to the nearest cluster in the dataset $x_i \in$ cluster j if $\|x_i - m_j\| < \|x_i - m_k\| \, k \neq j$.

3. Update the cluster centers m_j by calculating their average value, which is computed as x in cluster $j = \frac{1}{|c_l|} \sum_{x_i \in C_j} x, j = 1, \ldots k$.

4. Keep alternating 2 and 1 until m_j stops changing.

5. Repeat steps 2 and 3 until the cluster centers m_j stop changing, indicating that the algorithm has converged.

6. Ensure that neighboring pixels are in the same cluster by enforcing connectivity through a post-processing step that merges spatially close but unconnected clusters.

7. Finally, return the resulting clusters after completing the entire process.

EXTRACTION FEATURE

Extracting features from the super-pixels generated by the SLIC algorithm provides a means to capture distinctive attributes for each superpixel, outlining its defining characteristics. These feature descriptions encompass elements such as color, texture, and shape, offering valuable insights for image segmentation, object recognition, and scene understanding. By analyzing the features of individual super-pixels, researchers can gain a deeper understanding of the visual content within an image, allowing for more accurate and efficient analysis. Additionally, these feature descriptions are used to compare super-pixels across different images, enabling the development of algorithms that can recognize and classify objects based on their visual attributes. Gabor filters, on the other hand, utilize a set of spatial-frequency filters to capture the texture information present in each superpixel, allowing for more accurate texture-based analysis. Additionally, local binary patterns (LBP) can be used to encode the spatial arrangement of pixels within a superpixel, providing valuable shape information for further processing and analysis (Ren & Malik 2003).

As a result, combining relational analysis in conjunction with heterogeneous super-pixels and deep-fold features enables the identification of objects within images. Capsules demonstrate the feasibility of identifying superpixel attributes, highlighting the potential value of structured image analysis. The uniqueness of this deep learning architecture is tested in an image classification setting, with an emphasis on the network's explicable decision-making (Smith et al. 2010; Toth et al. 2018). Consider a collection of pixels denoted as V, which are partitioned into multiple disconnected segments that encompass all pixels within an image x. Each of these disconnected segments is associated with a combined value representative of region R. This process generates a superpixel representation within the feature space, as outlined by the Eq 8 (Yang et al. 2019):

$$y_i = \frac{\sum_{p \in R_j} x(p)}{|R_j|} \qquad (8)$$

Where p is a variable used for summation, and it represents the individual elements or pixels within a specific region R_i. Incorporating the concept of R, the representation accounts for how the different parts of the image connect as a whole. This restructuring of features aims to encompass substructures, thereby enhancing awareness of the overall structure of the image. Despite this rearrangement, the alignment between the feature space vectors and their positions within the image is preserved. Segmentation algorithms can be classified into four types: region-based, edge-constrained, classification or clustering, and hybrid approaches (Y. Chen et al. 2016).

In addition, k-means clustering can be applied to n-dimensional data sets, with the caveat that natural groups already exist. Natural clustering of n-dimensional data requires the user to choose a value of k. In addition, the super-pixels are color-segmented to classify the region of the virus's morphology. Iteratively adding neighboring pixels to cluster centers based on a similarity score yields this segmentation result. Common region-based strategies use region-growth algorithms, which necessitate the creation of comparable measurements and growth criteria. To classify unknown regions, these segmentation methods cluster them, and then classifiers are trained using the characteristics of individual super-pixels (Soltaninejad et al. 2017). The imsegkmeans was used for image segmentation based on the k-means clustering algorithm. It is generally used to divide images into regions or segments, with pixels in segments having similar colours, textures, or other characteristics. The imsegkmeans function takes an image and the number of clusters as input and returns an image where each pixel is assigned a label corresponding to the cluster. The function utilizes the k-means algorithm to find the cluster centres and assigns pixels to the closest cluster.

RESULTS AND DISCUSSION

Developing insight into the unique characteristics of the SARS-CoV-2 virus can shed light on the complex relationships between viruses and their hosts. Critical scientific phenomena like virus entrance, reproduction, mutation, escape mechanisms, viral abundance, and virus structure are all part of this web of interdependence.

Despite the complexities involved, many parallels exist between the techniques and methods of studying different viruses (Laue 2010). Figure 4 illustrates a manual analysis performed using ImageJ software to examine the morphological characteristics of the SARS-CoV-2 and SARS-CoV viruses. Specifically, Figure 4 (A) represents the original SARS-CoV-2 image with its segmented. In contrast, Figure 4 (B) shows the SARS-CoV authentic images with their segmented. The segmentation method was carried out so that the geometric parameters of each virus could be rigorously isolated and assessed, allowin for a thorough comparative study of their size and shape.

A. Original and Segmented Images of SARS-CoV-2

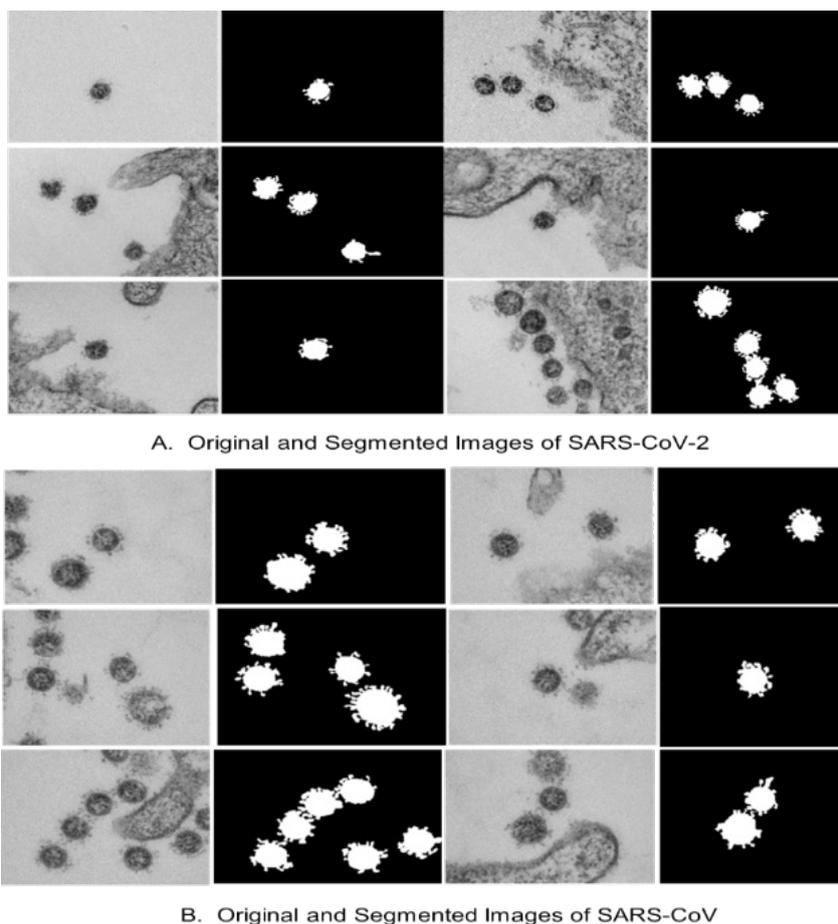B. Original and Segmented Images of SARS-CoV

FIGURE 4. Morphometry analysis of virus image using manually: A. Original and segmented images of SARS-CoV-2; B. Representing original and segmented images of SARS-CoV.

The unique characteristics of SARS-CoV-2 can aid in understanding how viruses interact with their hosts. Entrance, replication, mutation, escape mechanisms, virus levels, and structural features are all examples of viral interactions. These factors are subjects of significant interest in virus research. Nevertheless, it should be noted that the protocols for preparing and investigating various viruses share similarities (Laue 2010).

Figure 5 shows the results of the automatic detection of the SARS-CoV-2 and SARS-CoV area per pixel based on image segmentation with the superpixel algorithm. Several factors can explain variations in size for SARS-CoV-2 particles within the same image, such as the stage of maturation and/or activation. The location of the virus's replication cycle also impacts the particle density and size. Also, particles located at different positions within the cell may vary in size due to other environmental factors. Table 1 shows sample results to compare automatic and manual area segmentation matching.
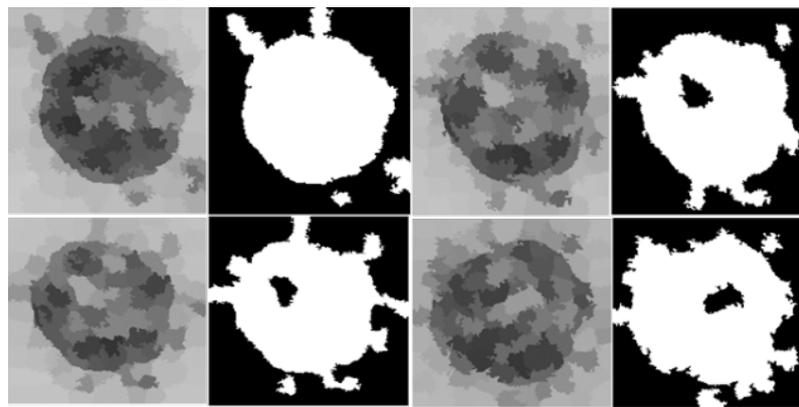
A. SARS-CoV-2 segmented
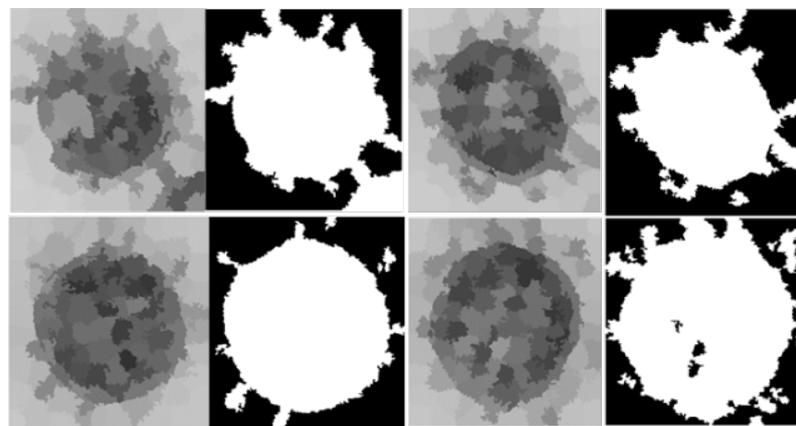
B. SARS-CoV segmented

FIGURE 5. illustrate the Auto-morphometry analysis of Coronavirus image using superpixel segmentation for A. SARS-CoV-2 and B. SARS-CoV

TABLE 1. Comparison of manual and auto area segmentation results for SARS-CoV-2 & SARS-CoV

Samples	Dataset images	Manually segmented	Auto segmented
1	Dataset_07_SARS-CoV-2_085	32325	31683
2	Dataset_02_SARS-CoV-2_088	25695	24786
3	Dataset_07_SARS-CoV-2_085	32325	31683
4	Dataset_03_SARS-CoV-2_092	22461	21991
5	Dataset_06_SARS-CoV_021	36220	36828
6	Dataset_06_SARS-CoV_016	32199	31627
7	Dataset_06_SARS-CoV_016	33916	34572
8	Dataset_06_SARS-CoV_101	34907	34177

Automated methods have become increasingly prevalent in various research fields, particularly in data analysis. The primary objective of using auto-methods is to achieve faster and more efficient data processing than traditional manual methods. Furthermore, the scalability of automated processes enables them to easily handle large datasets, which would be a daunting task if performed manually. Figure 6 shows the relationship between the value of the test samples and the mode of measurement used to determine the segmented area. The error values, calculated as the difference between the manual and automatic measurements, are plotted as individual points, providing a clear visual representation of the accuracy of the automated method.
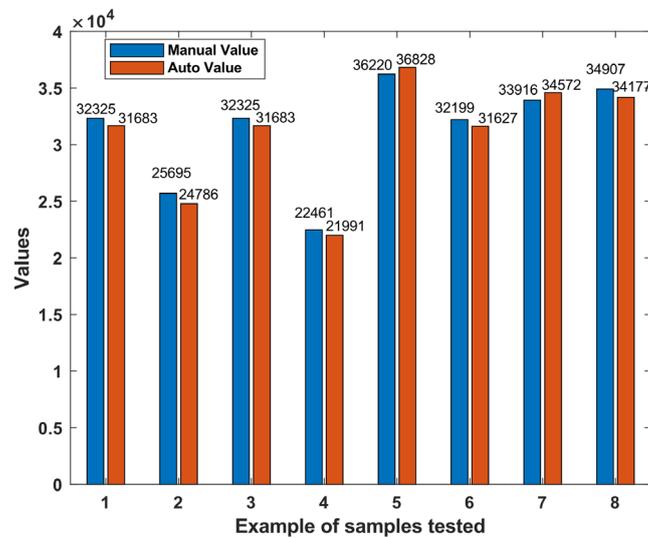
FIGURE 6. Morphometry analysis of virus image: A. Original and segmented images of SARS-CoV-2; B. Representing original and segmented images of SARS-CoV.
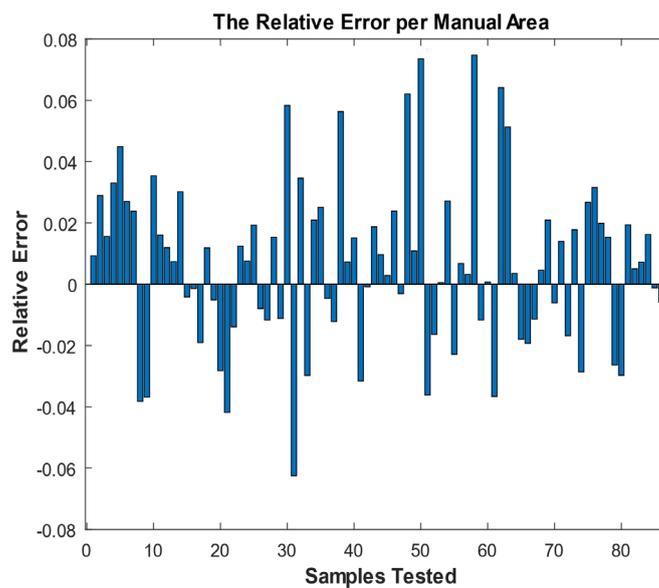

FIGURE 7. The error between the manual and Auto virus area segmentation

Figure 7 shows a discrepancy between the manual and automated measurements of the tested samples. The error determined was subtracting the automatic measurements from the manual ones. The results of the error analysis between the segmentation of the manual and automated virus area showed a low root mean square error (RMSE) of 0.0275, indicating a low level of variance between the two measurement methods. This relatively low error rate highlights the accuracy of the automated measurement method, suggesting an effective use in practical applications.

CONCLUSION

In conclusion, this paper has provided a new model for accurately and quickly detecting and classifying SARS-CoV-2 based on superpixel segmentation using TEM images. Furthermore, it includes valuable information on the mean area size of coronaviruses. This methodology offers useful information on the life cycle and progression of the virus, reducing the laborious manual work of experts. Additionally, this tool has significant potential for outbreak management and infection control, facilitating timely virus

diagnosis. It can also be extended to investigate other unknown viruses to improve early diagnosis and understanding of viral replication mechanisms and pathogenesis.

Furthermore, it represents a crucial step towards developing a robust deep learning model to accurately detect and classify SARS-CoV-2 levels in TEM images. Future work should consider incorporating more diverse data and fine-tuning the model parameters to enhance its performance. Finally, the findings of this study contribute to the advancement of automated virus detection in TEM images, empowering researchers and healthcare professionals with a powerful tool for efficient virus analysis, early detection, and improved understanding of viral diseases.

DECLARATION OF COMPETING INTEREST

None

REFERENCES

Abràmoff, M.D., Magalhães, P.J. & Ram, S.J. 2004. Image processing with imageJ. *Biophotonics International* 11(7): 36–41.

Abualigah, L.M., Khader, A.T. & Al-Betar, M.A. 2016. Multi-objectives-based text clustering technique using K-mean algorithm. *Proceedings - CSIT 2016: 2016 7th International Conference on Computer Science and Information Technology*: 12–17.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Susstrunk, S. 2010. SLIC Superpixels. *EPFL Technical Report 149300*, hlm.

Aggarwal, A., Mehta, S., Gupta, D., Sheikh, S., Pallagatti, S., Singh, R. & Singla, I. 2012. Clinical & immunological erythematosus patients characteristics in systemic lupus Maryam. *Journal of dental education* 76(11): 1532–9.

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. & Xia, L. 2020. Correlation of chest CT and RT-PCR testing for Coronavirus Disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology* 296(2): E32–E40.

Center for Systems Science and Engineering. 2021. COVID-19 Dashboard by the Center for Systems and Engineering at Johns Hopkins University.

Chen, Y., Zhang, Y., Yang, J., Cao, Q., Yang, G., Chen, J., Shu, H., Luo, L., Coatrieux, J.L. & Feng, Q. 2016. Curve-Like Structure Extraction Using Minimal Path Propagation with Backtracking. *IEEE Transactions on Image Processing* 25(2): 988–1003.

Chen, Z., Guo, B., Lib, C. & Liu, H. 2020. Review on Superpixel Generation Algorithms Based on Clustering. *Proceedings of 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education, ICISCAE 2020*: 532–537.

Colavita, F., Lapa, D., Carletti, F., Lalle, E., Bordi, L., Marsella, P., Nicastri, E., Bevilacqua, N., Giancola, M.L., Corpolongo, A., Ippolito, G., Capobianchi, M.R., Castilletti, C., Abbonizio, M.A., Agrati, C., Albarello, F., Amadei, G., Amendola, A., Antonini, M., Barbaro, R., Bartolini, B., Benigni, M., Bordoni, V., Branca, M., Campioni, P., Caporale, C., Caravella, I., Chiappini, R., Ciaralli, C., Cristofaro, M., Curiale, S., D'Abramo, A., Dantimi, C., de Angelis, A., de Angelis, G., Di Lorenzo, R., Di Stefano, F., Ferraro, F., Fiorentini, L., Frustaci, A., Gallì, P., Garotto, G., Giansante, F., Giombini, E., Greci, M.C., Lanini, S., Lepore, L., Lucia, A., Lufrani, F., Macchione, M., Marani, A., Marchioni, L., Mariano, A., Marini, M.C., Maritti, M., Matusali, G., Meschi, S., Messina, F., Montaldo, C., Murachelli, S., Nicastri, E., Noto, R., Palazzolo, C., Pallini, E., Passeri, V., Pelliccioni, F., Petrecchia, A., Petrone, A., Petrosillo, N., Pianura, E., Pisciotta, M., Pittalis, S., Proietti, C., Puro, V., Rinonapoli, G., Rueca, M., Sacchi, A., Sanasi, F., Santagata, C., Scarcia, S., Schininà, V., Scognamiglio, P., Scorzolini, L., Stazi, G., Vaia, F., Vairo, F. & Valli, M.B. 2020. SARS-CoV-2 isolation from ocular secretions of a patient with COVID-19 in Italy with prolonged viral RNA detection. *Annals of Internal Medicine* 173(3): 242–243.

Cong, J., Wei, B., Yin, Y., Xi, X. & Zheng, Y. 2014. Performance evaluation of simple linear iterative clustering algorithm on medical image processing. *Bio-Medical Materials and Engineering* 24(6): 3231–3238.

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P. & Ji, W. 2020. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 296(2): E115–E117.

De Geus, D., Meletis, P. & Dubbelman, G. 2019. Single

network panoptic segmentation for street scene understanding. *IEEE Intelligent Vehicles Symposium, Proceedings* 2019-June(Iv): 709–715.

Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P. & Johansen, H.D. 2020. DoubleU-Net: A deep convolutional neural network for medical image segmentation. *Proceedings - IEEE Symposium on Computer-Based Medical Systems* 2020-July: 558–564.

Joonas, L. 2020. Machine learning in safety critical industry domains Machine learning in safety critical industry domains.

Khan et.al. 2015. Image contrast enhancement using normalized histogram equalization. *Optik* 126(24): 4868–4875.

Kohler, P.F. & Farr, R.S. 1966. Elevation of Cord over maternal IgG immunoglobulin: Evidence for an active placental IgG transport [39]. *Nature* 210(5040): 1070–1071.

Kumar, V., Kandpal, D.C., Jain, M., Gangopadhyay, R. & Debnath, S. 2016. K-mean clustering based cooperative spectrum sensing in generalized κ-μ Fading channels. *2016 22nd National Conference on Communication, NCC 2016*: 1–11.

Kylberg, G., Uppström, M., Hedlund, K.O., Borgefors, G. & Sintorn, I.M. 2012. Segmentation of virus particle candidates in transmission electron microscopy images. *Journal of Microscopy* 245(2): 140–147.

Kylberg, G., Uppström, M. & Sintorn, I.M. 2011. Virus texture analysis using local binary patterns and radial density profiles. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7042 LNCS: 573–580.

Laue, M. 2010. Electron microscopy of viruses. *Methods in Cell Biology*, hlm. Elsevier Inc.:

Laue, M., Kauter, A., Hoffmann, T., Möller, L., Michel, J. & Nitsche, A. 2021. Morphometry of SARS-CoV and SARS-CoV-2 particles in ultrathin plastic sections of infected Vefile:///C:/Users/HP/Desktop/0645-06.pdf file:///C:/Users/HP/Desktop/nsmb1123.pdf ro cell cultures. *Scientific Reports* 11(1): 1–11.

Li, Shirahama, K. & Grzegorzek, M. 2015. Application of content-based image analysis to environmental microorganism classification. *Biocybernetics and Biomedical Engineering* 35(1): 10–21.

Luo, X. 2016. Image compression via K-means and SLIC superpixel approaches 71(Icmmita 2016): 1008–1012.

Maghdid, H.S., Asaad, A.T., Ghafoor, K.Z., Sadiq, A.S., Mirjalili, S. & Khan, M.K. 2021. Diagnosing COVID-19 pneumonia from x-ray and CT images using deep learning and transfer learning algorithms. *Proc.SPIE*, hlm. .

Matuszewski, D.J. & Sintorn, I.M. 2018. Minimal annotation training for segmentation of microscopy

images. *Proceedings - International Symposium on Biomedical Imaging* 2018-April(Isbi): 387–390.

Mettenleiter, T.C. 2017. The First "Virus Hunters." *Advances in Virus Research*, hlm. Edisi ke-1. Elsevier Inc.:

Ogando, N.S., Dalebout, T.J., Zevenhoven-Dobbe, J.C., Limpens, R.W.A.L., van der Meer, Y., Caly, L., Druce, J., de Vries, J.J.C., Kikkert, M., Barcena, M., Sidorov, I. & Snijder, E.J. 2020. SARS-coronavirus-2 replication in Vero E6 cells: Replication kinetics, rapid adaptation and cytopathology. *Journal of General Virology* 101(9): 925–940.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J. & Qian, Z. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications* 11(1).

Ren, X. & Malik, J. 2003. Learning a classification model for segmentation. *Proceedings of the IEEE International Conference on Computer Vision* 1(c): 10–17.

Richert-Pöggeler, K.R., Franzke, K., Hipp, K. & Kleespies, R.G. 2019. Electron microscopy methods for virus diagnosis and high resolution analysis of viruses. *Frontiers in Microbiology* 10(JAN): 1–8.

Rivenson, Y., Wu, Y. & Ozcan, A. 2019. Deep learning in holography and coherent imaging. *Light: Science and Applications* 8(1).

Robertson, S., Azizpour, H., Smith, K. & Hartman, J. 2018. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Translational Research* 194: 19–35.

Roingeard, P., Raynal, P.I., Eymieux, S. & Blanchard, E. 2019. Virus detection by transmission electron microscopy: Still useful for diagnosis and a plus for biosafety. *Reviews in Medical Virology* 29(1): 1–9.

Sahiner, B., Pezeshk, A., Hadjiiski, L.M., Wang, X., Drukker, K., Cha, K.H., Summers, R.M. & Giger, M.L. 2019. Deep learning in medical imaging and radiation therapy. *Medical Physics* 46(1): e1–e36.

Schwartz, D.A. 2020. An analysis of 38 pregnant women with COVID-19, their newborn infants, and maternal-fetal transmission of SARS-CoV-2: Maternal coronavirus infections and pregnancy outcomes. *Archives of Pathology and Laboratory Medicine* 144(7): 799–805.

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y. & Shen, D. 2021. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering* 14(c): 4–15.

Siam, M., Elkerdawy, S., Jagersand, M. & Yogamani, S.

2018. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2018-March: 1–8.

Smith, K., Achanta, R., Knott, G. & Fua, P. 2010. Supervoxel-Based Segmentation of EM Image Stacks with Learned Shape Features. *Scanning Electron Microscopy* 31(December): 1–12.

Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T.L., Barrick, T.R., Howe, F.A. & Ye, X. 2017. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *International Journal of Computer Assisted Radiology and Surgery* 12(2): 183–203.

Spilger, R., Imle, A., Lee, J.Y., Muller, B., Fackler, O.T., Bartenschlager, R. & Rohr, K. 2020. A Recurrent Neural Network for Particle Tracking in Microscopy Images Using Future Information, Track Hypotheses, and Multiple Detections. *IEEE Transactions on Image Processing* 29: 3681–3694.

T. Ferreira, W.R. 2012. ImageJ User Guide. *IJ 1.46r*: 185.

Taha, B.A. 2021. Perspectives of Photonics Technology to Diagnosis COVID–19 Viruses: A Short Review. *Journal of Applied Sciences and Nanotechnology* 1(1): 1–6.

Taha, B.A., Ali, N., Sapiee, N.M., Fadhel, M.M., Mat Yeh, R.M., Bachok, N.N., Al Mashhadany, Y. & Arsad, N. 2021a. Comprehensive Review Tapered Optical Fiber Configurations for Sensing Application: Trend and Challenges. *Biosensors* 11(8): 253.

Taha, B.A., Al-Jubouri, Q., Al Mashhadany, Y., Hafiz Mokhtar, M.H., Bin Zan, M.S.D., Bakar, A.A.A. & Arsad, N. 2023a. Density estimation of SARS-CoV2 spike proteins using super pixels segmentation technique. *Applied Soft Computing* 138: 110210.

Taha, B.A., Al-Jubouri, Q., Al Mashhadany, Y., Zan, M.S.D. Bin, Bakar, A.A.A., Fadhel, M.M. & Arsad, N. 2022a. Photonics enabled intelligence system to identify SARS-CoV 2 mutations. *Applied Microbiology and Biotechnology* 106(9): 3321–3336.

Taha, B.A., Al Mashhadany, Y., Al-Jubouri, Q., Rashid, A.R.B.A., Luo, Y., Chen, Z., Rustagi, S., Chaudhary, V. & Arsad, N. 2023b. Next-generation nanophotonic-enabled biosensors for intelligent diagnosis of SARS-CoV-2 variants. *Science of The Total Environment* 880: 163333.

Taha, B.A., Mashhadany, Y. Al, Al-Jumaily, A.H.J., Zan, M.S.D. Bin & Arsad, N. 2022b. SARS-CoV-2 Morphometry Analysis and Prediction of Real Virus Levels Based on Full Recurrent Neural Network Using TEM Images. *Viruses* 14(11): 2386.

Taha, B.A., Al Mashhadany, Y., Bachok, N.N., Ashrif A Bakar, A., Hafiz Mokhtar, M.H., Dzulkefly Bin Zan, M.S. & Arsad, N. 2021b. Detection of covid-19 virus on surfaces using photonics: Challenges and perspectives. *Diagnostics* 11(6): 1119.

Taha, B.A., Mashhadany, Y. Al, Mokhtar, M.H.H., Zan, M.S.D. Bin & Arsad, N. 2020. An analysis review of detection coronavirus disease 2019 (Covid-19) based on biosensor application. *Sensors (Switzerland)* 20(23): 1–29.

Taha, B.A., Mehde, M.S., Haider, A.J. & Arsad, N. 2022c. Mathematical model of the DBR laser for thermal tuning: taxonomy and performance effectiveness with PbSe materials. *Journal of Optics*.

Thiel, V., Ivanov, K.A., Putics, Á., Hertzig, T., Schelle, B., Bayer, S., Weißbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., Gorbalenya, A.E. & Ziebuhr, J. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology* 84(9): 2305–2315.

Toth, T., Balassa, T., Bara, N., Kovacs, F., Kriston, A., Molnar, C., Haracska, L., Sukosd, F. & Horvath, P. 2018. Environmental properties of cells improve machine learning-based phenotype recognition accuracy. *Scientific Reports* 8(1): 1–9.

Vidyasaraswathi, et al. 2020. Low-Contrast Image Enhancement Histogram method using Grey Wolf Optimization search algorithm 15(1): 19–28.

Yang, A., Veal, C.T., Anderson, D.T. & Scott, G.J. 2019. Recognizing Image Objects by Relational Analysis Using Heterogeneous Superpixels and Deep Convolutional Features: 1–10.

Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., Shi, Y., Li, L. & Li, S. 2020. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* 183(3): 730-738.e13.

Yuan, Y., Zhang, W., Yu, H. & Zhu, Z. 2021. Superpixels with Content-Adaptive Criteria. *IEEE Transactions on Image Processing* 30: 7702–7716.

Zhao, J., Zhou, H., Huang, W., Zhou, J., Qiu, M., Deng, Z., Chen, L., Weng, Y., Cai, L., Gu, Y., Zheng, Q., Chen, Q., Hou, X., Wang, L., Shen, L. & Yang, Z. 2020. Cell morphological analysis of SARS-CoV-2 infection by transmission electron microscopy. *Journal of Thoracic Disease* 12(8): 4368–4373.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F. & Tan, W. 2020a. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* 382(8): 727–733.

Zhu, Z., Lian, X., Su, X., Wu, W., Marraro, G.A. & Zeng, Y. 2020b. From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respiratory Research* 21(1): 1–14.