They have proposed an automatic end-to-end deep neural network algorithm for segmentation of individual nuclei. A nucleus-boundary model is introduced to predict nuclei and their boundaries simultaneously using a fully convolutional neural network [19].

nuceli segmentation:

This paper addresses the task of nuclei segmentation in high-resolution histopathological images. We propose an automatic end-to-end deep neural network algorithm for segmentation of individual nuclei. A nucleus-boundary model is introduced to predict nuclei and their boundaries simultaneously using a fully convolutional neural network. Given a color normalized image, the model directly outputs an estimated nuclei map and a boundary map. A simple, fast and parameter-free post-processing procedure is performed on the estimated nuclei map to produce the final segmented nuclei. An overlapped patch extraction and assembling method is also designed for seamless prediction of nuclei in large whole-slide images. We also show the effectiveness of data augmentation methods for nuclei segmentation task. Our experiments showed our method outperforms prior state-of-the art methods. Moreover, it is efficient that one 1000X1000 image can be segmented in less than 5 seconds. This makes it possible to precisely segment the whole-slide image in acceptable time.

This study presents a comparative study of twelve nuclei segmentation methods for cytology pleural effusion images. Each method involves three main steps: preprocessing, segmentation, and postprocessing. preprocessing and segmentation stages help enhancing the image quality and extracting the nuclei regions from the rest of the image, respectively. postprocessing stage helps in refining the segmented nuclei and removing false findings. segmentation methods are quantitatively evaluated for 35 cytology images of pleural effusion by computing five performance metrics. Evaluation results show that the segmentation performances of the Otsu, k-means, mean shift, ChanVese, and graph cut methods are 94, 94, 95, 94, and 93%, respectively, with high abnormal nuclei detection rates [20].

In this study, to solve the segmentation of nuclei and overlapping regions, we introduce a nuclei segmentation method based on two-stage learning framework consisting of two connected Stacked U-Nets (SUNets) [21].

image processing technique such as thresholding and contour based. The segmented nucleus present in the histopathological images are classified as benign and malignant. The histopathological images are made up of different structures such as the nucleus, chromosomes, cytoplasm, lymphocytes etc.

The methods proposed and work detailed in this paper have the following key highlights

- **An Xception-style UNet with skip connections is experimented and proposed to segment nuclei from a Whole Slide Histopathological Image (WSHI) patch. The network achieves a mean-IoU of 0.9427 on the test data.**
- **An algorithmic method of generating masks as ground-truth for training and testing the UNet network is proposed. The method is based on contour detection and color-based filtering techniques.**
- **A Deep Convolutional Neural Network (DCNN) is proposed to perform binary and multi-class classification of lung tumors using complete WSHI patches as well as segmented nuclei information obtained through the proposed deep learning segmentation network.**

3. OVERVIEW OF PROPOSED SYSTEM

In this paper, we propose a Deep Convolutional Neural Network (DCNN) that performs *two forms* of classification of the histopathological lung images — binary and multi-class. While binary classification identifies an input image as either benign or malignant, multi-class classification will assign one of three classes to an input image — ADC, SCC and Benign.

The nuclei of benign and cancerous lung tissues are known to posses visually distinguishable characteristics. Furthermore, each sub-type of lung cancer affects these characteristics differently. Limiting the information available in the input images to only these relevant regions that impact classification, can improve the classifier's performance. A DCNN classifier extracts features from the information available to it. When only nuclear regions are available, the effort of the DCNN in learning to suppress the weights of less relevant features stemming from other regions of the image is greatly reduced. Moreover, retaining only the nuclear regions from a WSHI patch gives rise to images that mask the non-nuclear regions with sharp gradients at the nucleus boundaries. As a result of such highly pronounced boundaries, shape-based features will be learnt much more robustly and rapidly.

Hence, we propose to improve the classification accuracy of the DCNN by performing semantic segmentation to retain only the nuclear regions of the input WSHI patches. We present an Xception-style UNet for semantic segmentation, along with *two different approaches* to prepare the ground-truth for training this network. Further, we study and present the difference in performance of the classifier when it is trained with input images that contain only the segmented nuclear regions from when the entire WSHI patches are employed.

3.1. Preliminiary Preprocessing

Our proposed methodology involves training a DCNN to perform classification. To avoid over fitting, to

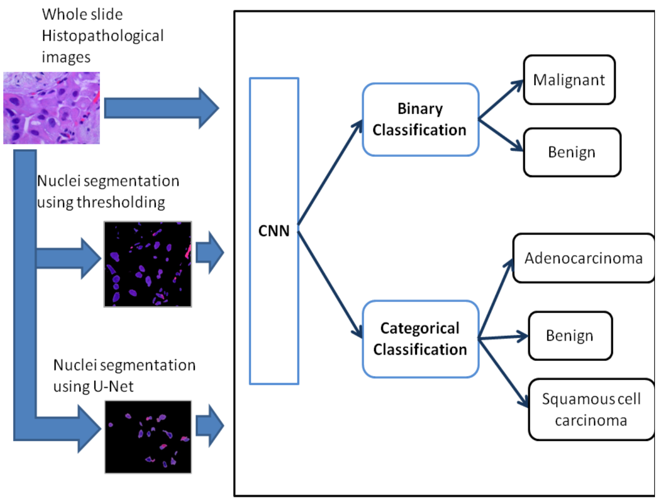

FIGURE 2. Proposed system

adopt a diversity-based sampling strategy and to cater to the large input requirement of the deep learning network, data augmentation is necessary. Random affine transformations such as rotation, scaling and shearing are applied to augment the input data. The input images are resized to 128x128, 3-channel images using bi-linear interpolation. The images are then linearly-normalized to envelope the pixel values between 0 and 1. This is known to improve the rate of convergence and enhance the network's ability to standardize to diverse inputs.

3.2. Input Image Preparation

After the preliminary stage of preprocessing, two different pipelines are adopted — the first, feeds the preprocessed WSHI patches to the classifier directly and the second, extracts only the nuclear regions of the input images and passes them to the classifier.

The second pipeline is of particular interest, which involves performing an automated semantic segmentation of the preprocessed WSHI patches to retain only the nuclear information from the image. This is achieved with the help of a segmentation network which takes an input WSHI patch and produces a binary mask image. A binary mask is an image with the same dimensions as the input image, but the pixel values are replaced with a numeric class label to classify the corresponding pixels in the input image into one of two semantic class — namely, nuclear and non-nuclear regions. To this effect, the mask encodes nuclear regions with a class label of 1 and 0 otherwise.

A *bitwise-AND* operation is performed between the input RGB image of the WSHI patch and the generated mask. The result of this operation is an image containing only the information corresponding to the nuclear regions of the input WSHI patch as encoded in the mask. This operation is depicted in the Figure 3. In the second pipeline, the image so obtained is passed to the DCNN classifier. The specific details about the segmentation network and the methods adopted to prepare the ground truth for training the network are included in a subsequent section (section-4).

3.3. Binary and Multi-Class Classification

The input WSHI patch images, after being processed through one of the pipelines, is relayed to a DCNN classifier. Two variants of the classifier are proposed — one for binary classification and the other for multi-class classification. The binary classifier labels the input WSHI patch as either benign or malignant, and the multi-class variant goes one step further and assigns one of three classes to an input image — ADC, SCC and Benign, where ADC and SCC are sub-classes of the malignant type.

The two variants of the classifier are kept consistent across all the pipelines to evaluate and contrast between the quality of classification achieved through each of the different input preparation methods and segmentation approaches adopted. The classifier network architecture and its variants as well as an analysis of the results obtained through each method are detailed in separate subsequent sections (section-5 and section-6).

4. SEMANTIC SEGMENTATION TO EXTRACT NUCLEAR REGIONS

The second pipeline proposed for classification involves extracting nuclear regions of the input WSHI patches through semantic segmentation. We propose an Xception-style UNet architecture to perform this binary segmentation. The remainder of this section provides some background on the rationale behind the methods adopted to prepare the ground-truth for semantic segmentation, the preparation procedure and details of the network architecture used in this work.

4.1. Ground Truth Preparation

During a histopathological examination, the obtained tissue samples are prepared on microscope slides for visual examination, typically using methods such as chemical fixation and frozen section processing. Such slides are stained using coloring pigments to reveal and distinguish cellular components. Hematoxylin and Eosin (H&E) is the most commonly used stain combination. Hematoxylin reacts like a basic dye with specific cell components — namely the cell nucleus, ribosomes and the endoplasmic reticulum. The basic reaction stains these cell components with a purplish-blue color. On the other hand, Eosin is an acidic dye that affixes the cytoplasm, cell walls and extracellular connective matrix of the tissue with a dull reddish-pink stain. These slides can then be examined under a microscope by pathologists or micrographed to produce WSHIs.

The WSHI patch images used in this study are stained using H&E. On account of staining, the nuclear regions (as well as some other cellular components) in the histopathological images have significantly different visual properties than the remainder. The visual properties of any image pixel can be broadly categorized into two — *chroma* and *luma*. While *chroma* represents the color of a pixel, *luma* encodes the intensity level of that color in the pixel. Based on this categorization, two different approaches to extracting the nuclear regions from the WSHI patches can be conceived — one based on intensity variations and the other, rooted at color distinctions. These approaches are pursued to prepare the ground-truth masks required to train the segmentation network.

It is worth noting, that while Hematoxylin stains all the nuclear regions of the tissue, it stains a few other cellular components as well — namely, the ribosomes and the endoplasmic reticulum. Hence, a segmentation procedure that obtains the nuclear regions based on the variations introduced by staining, will inevitably include some portions of the non-nuclear regions as well. These portions are regarded as noise. However, the intention behind the approach proposed in the second pipeline is to limit the image regions available to the classifier to contain only the most relevant information. Hence, as long as none of the nuclear regions of the image are lost to segmentation, allowing some non-nuclear regions of the image to persist in the segmented input is not detrimental. It will not degrade the classification performance any more than passing the entire WSHI patch, as is the case in the first pipeline. In the remainder of this paper, the mention of nuclear regions in the context of segmentation, implicitly includes the noise that may accompany it.

4.1.1. Based on Intensity Variation

This approach to segmenting the nuclear information from input WSHI patches leverages the intensity variations in the visual structure of the WSHI patches due to staining. The nature of staining leaves the pixels of the nuclear regions with a darker shade and hence a higher intensity value, relative to the remainder of the image.

The input patch images are first converted to grayscale to retain only the intensity information of pixels [22] and discard the color details. Otsu's method [23], a global thresholding method that determines an optimal threshold to segregate the pixels in a given image into two classes with minimum intra-class variance and maximum inter-class differentiation, is then applied to obtain an image-global intensity threshold value. In effect, these two classes of pixels correspond to the nuclear and non-nuclear regions of the WSHI patch, with the higher intensity class representing the former.

The pixels with intensity values lower than the obtained threshold are folded down to a value of 0 while the remainder are set to 1, thereby generating a binary image. Semantically, this image retains an extract of the WSHI patch that contains only the nuclear regions of the image, as identified by the thresholding technique, by encoding those pixels with an intensity value of 1. The binary image is then extrapolated into a three-channel image by duplicating the only channel present. This serves as a binary mask that can be used to extract nuclear information from the corresponding WSHI patch by performing a *bitwise-AND* operation between the two.

The WSHI patch image and the corresponding binary mask serve as input and annotation respectively, to train the segmentation network.

4.1.2. Based on Color Distinction

Staining histopathological slides with HE, attributes a purplish-blue color to the nuclear regions of the image. In this approach, nuclear information is extracted from input WSHI patches by exploiting this color distinction as well as the intensity variations that arise in the visual structure of the WSHI patches on account of staining.

The input WSHI patch is first partitioned into several sub-regions. A sub-region here, refers to the set of pixels enclosed within each boundary representing a subset the image area. This is done using Suzuki's contour detection algorithm [24] and it uses the variation in intensity of pixels to achieve this. These variations are more pronounced at the boundary between nuclear and non-nuclear regions of the image and as a result, transitioning from a nuclear component of the image to an adjacent non-nuclear region is characterized by a sharp change in the pixel intensity values. These abrupt variations are detected by the contour detection algorithm to find closed loop boundaries within the image, that separate adjacent nuclear and non-nuclear components. The boundaries so obtained are stored as a vector of continuous pixels forming a closed loop and can be applied on the WSHI patch to obtain sub-regions. In effect, the image is partitioned into a large number of sub-regions, some pertaining to nuclear regions and the others representing non-nuclear components in the patch.

The WSHI patch with the obtained sub-regions is a RGB image. RGB images encode *chroma* and *luma* information in an inter-mixed manner. The Hue-Saturation-Value (HSV) image format on the flip side, separates the two, with the hue component encoding all the *chroma* information and the Saturation and Value components containing the *luma* information about color saturation and brightness, respectively.

In order to use only the color information, the WSHI patch is now converted to the HSV format. For each sub-region in the patch image, the average hue component value is computed. A suitable range of average hue values that can describe the nuclear sub-regions is arrived at, using expert guidance. Finally, the sub-regions having average hue values within this range are retained in the WSHI patch image, discarding the rest. These sub-regions are then used to produce a corresponding binary mask image with nuclear regions encoded with a pixel value of 1 and the non-nuclear regions i.e the regions discarded from the patch image, marked as 0. The resulting mask image, therefore, serves as an extract of nuclear information.

When determining the hue value range, a trade-off must be made between two options that arise from the inherent approximation associated with the method used to find the sub-regions. On one hand, the segmented image can be restricted to disallow any component apart from the nucleus at the expense of losing some nuclear information. On the other hand, it can be ensured that the segmented image has all the nuclear information at the cost of allowing some portions of other components. The latter is chosen, since feeding some extra information to the classifier DCNN, while guaranteeing the presence of all nuclear information, is a more justified approach than compromising important nuclear information.

As in the previous approach, the WSHI patch image and the corresponding binary mask so obtained, serve as input and annotation respectively, to train the segmentation network.

4.2. Segmentation Network Architecture

The UNet [25] is a Convolutional Neural Network architecture, proven for segmentation of medical images. The segmentation of nuclei information from the WSHI requires just two semantic classes — nuclear and non-nuclear regions. An initial experiment with a simple UNet architecture showed accuracy saturation and high training loss, suggestive of the degradation problem [26]. Hence, we propose an Xception-style [27] UNet architecture, to employ residual blocks that can reduce degradation in deep networks and easily adapt to learning simplex as well as complex correlations between the layers of the network [28]. Concurrent with the standard UNet structure, the proposed architecture consists of a contractive downsampling path followed by an expansive upsampling path. The input images are first resized to 512×512 with 3 channels and passed through an entry block. This block comprises of a convolution layer with a standard kernel size of 3×3, a batch-normalization layer and a Rectified Linear Unit (ReLU) activation layer. The block produces an output with 32 channels.

The contracting path is composed of three blocks with progressively increasing filter sizes of 64, 128 and 256. The intention is to learn about the existence of distinctive finer features as the image percolates deeper into the network. Separable Convolution layers are used, which perform depth-wise spatial convolutions followed by point-wise convolutions that mix the output channels. The approach is adopted with the supporting hypothesis on Xception networks that cross-channel correlations and spatial correlations can be treated as being entirely disengaged. In each block, two sets of convolution layers, each preceded by a ReLU activation layer and followed by a batch normalization layer is applied. This sequence is followed by a max-pooling layer and an addition-based skip connection from the input to the current block. The placement and ordering of the layers within each block determines the nature of activation. As discussed in [26], pre-activation and post-activation produce significantly distinct network performances in the presence of element-wise addition. Element-wise addition is introduced by the use of residual connections. Specifically, pre-activation produces better regularization by reducing overfitting and eases optimization due to a more direct weight propagation between subsequent blocks, when the relationship between them is more closer to an identity mapping rather than a complex function. The ReLU-only pre-activation and the full pre-activation approaches discussed in [26] were experimented with, and the former approach was found to produce better results. Concretely, the sequence shown in sequence-(3) is adopted within each block of the downsampling strata.

$$SkipConnectionStartPoint \rightarrow ReLU \rightarrow Convolution \rightarrow BatchN \tag{1}$$

The latter half of the network is an expanding path that uses transposed convolutions [29] to upsample the low resolution image at the end of the contracting path progressively, back to its original resolution. A single step of transposed convolution is achieved by performing convolutions on the image obtained by sufficiently zero-padding around each pixel of the low resolution image. In effect, the subsequent series of downsampling and upsampling operations, recovers the spatial information lost during the contraction phase and further, localizes the identified features to their corresponding positions on the higher resolution image produced at each step of the expansion path. Ultimately, pixel-wise classification into semantic classes is perfomed on these localized features once the original resolution is reached. To further enhance the precision of mapping between features and their respective locations on the upsampled image and avoid losing fine-grained information, skip connections are used to forward and concatenate features from corresponding stages of the contraction path with the upsampled image obtained at the same stage of expansion path. In particular, the use of skip connections is known to help with feature reuse and is the most notable purpose of employing them in the segmentation network. However, shorter skip connections are also known to alleviate the vanishing gradient problem [30, 31].

The transpose convolution blocks used in the expansion path have a symmetric structure with respect to the contracting path. It comprised of three blocks with progressively decreasing filter sizes of 256, 128 and 64. Following the symmetric structure, each block is constructed with two sets of transpose convolution layers, each followed by a ReLU activation layer and followed by a batch normalization layer. This is followed by an upsampling layer and an addition-based skip connection to forward features from the current block's input to output. Once again, the sequence of layers in each block follows the ReLU-only pre-activation strategy. Notably, the upsampling layer is different from a convolution transpose layer in that the former simply doubles the dimensions of its input while the latter performs an inverse convolution operation and learns the kernel during the training phase. In particular, each block of the upsampling strate follows the sequence depicted in sequence-(2)

$$SkipConnectionStartPoint \rightarrow ReLU \rightarrow Convolution \rightarrow BatchNormalization \rightarrow ReLU \rightarrow Convolution \rightarrow BatchNormaliza \tag{2}$$

Two variants of this network are trained using complete WSHI patches and their corresponding annotation masks obtained using the two ground truth preparation approaches. When the second pipeline is adopted for classification, the input WSHI patches are passed through each of these variants to obtain segmentation masks. The *bitwise-AND* operation is performed between the mask and the input patch image to produce an image containing the extract of only the nuclear regions. Finally, this extract image is passed to the DCNN classifier. The quality of segmentation obtained from each variant of the network as well as the final classification performance resulting from each of the ground truth preparation approaches is compared and analyzed in a subsequent section (section-6).

5. CLASSIFICATION OF LUNG TUMOR CELLS

5.0.1. Complete WSHI Patches

In this approach, the input images obtained after data augmentation are directly fed to DCNN for classification. Consequently, the features extracted by the network for classification are not limited to the nuclear regions, but span throughout the input WSHI patch. This is the most commonly adopted strategy in histopathological image classifiers. We have included this method in our work to establish a baseline performance measure of the proposed DCNN classifier. This baseline is used to study and discriminate the impact of segmenting out nuclear features from the WSHI patches on the accuracy of classification.

5.0.2. Intensity Based Thresholding to Segment the Nuclear Information

This approach to segmenting the nuclear information from WSHI patches utilized the intensity variations of n

The WSHI patch images used in this study are stained using HE. Hence the nuclei and Since different components of the histopathological images are stained differently, we propose to segment the nucleui from the WSHI based on the color variations. Due to the nature of staining, the nuclei is always of a brighter shade in the grayscale images. This characteristic is exploited to segment the nuclei from the WSHI.

Image pixels are charecterized by color and intensity

images are converted into a gray-scale image. Otsu's method, a global thresholding method that determines an optimal threshold for the given imagethat minimizes the intra-class variance between the nucleus and non-nucleus sections of the image, is used to obtain a threshold. The pixel values lower than the obtained threshold are set to 0, while the others are set to 1. The output image contains segmented nucleus in the binary form along with some noise, stemming from other stained parts of the cells. The output is converted into a 3-channel image by duplicating the only channel present and a bit-wise 'and' operation is performed with the RGB input image. This extracts and retains only the information corresponding to the regions identified as nuclear information by the thresholding technique, from the WSHI. This operation is depicted in the Figure 3

We propose a CNN model comprising of six convolution blocks. Each block contains a convolution layer, an activation layer, a batch-normalization layer and a max-pooling layer. Each block consists of a convolution followed by activation layer, a batch normalization layer and a pooling layer. Dropout layers are added at the end of alternate convolution blocks to prevent over-fitting. The convolution blocks extract distinctive features from the input images. The initial layers extract high-level features. As the images percolate through the deeper layers of the network, the channel-size is increased progressively from 32 to 128 to extract more subtle features that can help in inter-class differentiation. The feature representation vectors obtained from the convolution layers are flattened using the flatten layer and fed into the classification layer. The classification layer consists of two dense stages. The first stage includes a dense layer followed by an activation layer and a dropout layer. The second stage, which is the output layer, includes a dense layer followed by an activation layer. The activation layer consist of a single neuron for binary classification and three neurons for multi-