# A comparative study of image feature descriptors to classify leukemia cancer cells from microscopic images using deep learning

Manisha Suryakant Takale
STUDENT NUMBER: 2032988

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

dr. Juan Sebastian Olier Jauregui
dr. Grzegorz Chrupala

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
June 2020

**Preface**

This study was conducted under the supervision of dr. Juan Sebastian Olier Jauregui. I would like to express my sincere gratitude and appreciation to my supervisor whose guidance, support and encouragement has been very helpful throughout this research.

# A comparative study of image feature descriptors to classify leukemia cancer cells from microscopic images using deep learning

Manisha Suryakant Takale

*The numerous computer-aided systems use microscopic images to identify acute lymphoblastic leukemia (ALL). The general approach for automatic detection of ALL is to use transfer learning features with pre-trained models such as AlexNet, VGG16, MobileNet, Inception-v3. However, various image feature descriptors like local binary patterns (LBP), speeded up robust features (SURF), the scale-invariant feature transform (SIFT), and the histogram of oriented gradient (HOG) efficiently represents the essential information in the image. In this paper, a comparative study of these feature descriptors carried out for the automatic detection of leukemia. Furthermore, a shallow convolutional neural network (S-CNN) with six-layers and 2M trainable parameters implemented to train various feature descriptors. From the comparative analysis, the proposed deep learning model efficiently identifies cancer cells from microscopic images with an accuracy of 94.59% on original images and 91.04% on HUE-SIFT images. The final experimental results show that S-CNN works best on pre-processed original images compare to feature descriptors used in this paper to detect leukemia cells from microscopic images.*

## 1. Introduction

White blood cells (WBC's) and red blood cells (RBC's) are two major components of blood along with plasma and platelets. White blood cells play a significant role in the normal functioning of the human immune system. However, sometimes irregular functioning of the bone marrow (the soft tissues present inside the bone), causes rapid production of immature WBC's. This abnormal production of WBC's is called leukemia cancer (Rawat et al. 2017b). The four major types of leukemia are summarized in Table 1. Out of these four types, Acute Lymphocytic leukemia (ALL) is considered fatal because it weakens the immune system and makes the human body vulnerable to diseases (Thanh et al. 2018).

The word "acute" in Acute Lymphocytic leukemia means a rapid increase in immature blood cells rather than mature ones, and the word "Lymphocytic" means white blood cells named lymphocytes present in the blood. The abnormal production of immature blood cells first affects the immune system and eventually spreads to other organs such as the liver, spleen, and lymph nodes. ALL is the most common adolescent cancer, with an incident rate of 19.8% per 100000 children between the age group of 0 to 19 (LaRussaA 2015). However, early detection of leukemia can help with early treatment and cure. Even though ALL is not very common in adults, it may be fatal and difficult to cure if diagnosed as positive. This study focuses on differentiating malignant blood cells from healthy cells of the B-ALL type of cancer using microscopic images.

The abnormal WBC's are immature blood cells with a larger nucleus and are low in the cytoplasm. Therefore, the microscopic images of blood consist of information such

**Table 1**
Four main types of leukemia

| | **Acute** | **Chronic** |
|---|---|---|
| **Lyphocytic** | Acute lymphoblastic leukemia (ALL) | Chronic lymphocytic leukemia (CLL) |
| **Myelogenous** | Acute myeloid leukemia (AML) | Chronic myeloid leukemia (CML) |

as shape, texture, orientation, and color space. Thus, various feature descriptors can extract vital information from images to differentiate healthy cells from cancer cells. This paper investigates four widely used feature descriptors such as local binary patterns (LBP), speeded up robust features (SURF), scale-invariant feature transform (SIFT), the histogram of oriented gradient (HOG) for leukemia cell detection in an image.

### 1.1 Motivation

This study's primary motivation is to use feature descriptors LBP, SURF, SIFT, HOG, and train them with a shallow convolutional neural network (S-CNN) to identify leukemia blood cells. The texture information present in microscopic images can be effectively presented in the grayscale image using the LBP algorithm. LBP is a widely used texture feature descriptor that is highly discriminative and computationally efficient (Pietikäinen and Zhao 2015). The SURF and SIFT algorithms effectively extract local information that helps to detect objects in an image. These algorithms are invariant to scale, rotation, blur, illumination, and noise. Hence, similar objects in different images can be matched together even though images vary in terms of scale, rotation, blur, illumination, and noise (Kashif et al. 2016). The shape of the objects present in an image can effectively be separated based on its local gradients and edge direction. When applied to the microscopic image, the HOG algorithm can generate a grayscale image that clearly distinguishes the core shape of the blood nucleus, and these shapes can be represented using local gradients.

A literature review has shown that most of the microscopic image classification uses either transfer learning or segmentation for feature extraction. The transfer learning technique mainly uses various pre-trained deep learning models such as VGG16, ResNet50, AlexNet, Inception-v3, and MobileNet to extract features from the image (Simonyan and Zisserman 2014; He et al. 2016; Krizhevsky, Sutskever, and Hinton 2012; Szegedy et al. 2016; Howard et al. 2017). However, these pre-trained models have few limitations, such as they are big in terms of architecture and consist of a large number of trainable parameters. In general, these trainable parameters range from 60M (AlexNet) to 140M(VGG16) (Teuwen and Mariakav 2020) depending upon the size of input image. Another drawback of these models is that they are designed with many numbers of layers and big in storage size (from 50MB to 500MB) (Cui et al. 2019). Furthermore, these heavy models are trained on natural images rather than microscopic images and require massive hardware setup for training.

Therefore, this study also focuses on implementing a shallow convolutional neural network (S-CNN) corresponding to the state-of-the-art pre-trained neural networks. Compare to more complex networks, this research will have the architecture of choice

with a single convolutional layer per block and less number of filters. Furthermore, the proposed S-CNN will have a smaller number of layers and trainable parameters. Due to the simplicity of the model, one can train it using simple hardware setup or on more data for better generalization.

### 1.2 Research Question

The research question that will be answered in this paper is **To what extent can leukemia blood cell be identified using a shallow convolutional neural network trained on different feature descriptors?**

Along with the main research question, this study will explore below sub-questions:

1. Which feature descriptor yields better performance with the proposed shallow net, and why?
2. Which technique, a shallow net with feature descriptor or VGG16 with transfer learning, produces the best classification accuracy?

For the comparative analysis of feature descriptors on microscopic images, this study uses CNMC_B-ALL dataset released in 2019 by SBILab. Since the dataset is released recently, very few automatic detection systems exist to detect leukemia using this dataset. The main difference between the previously released leukemia datasets and CNMC_B-ALL dataset is that each microscopic image consists of a single blood cell; however, the prior dataset consists of multiple blood cells per image. Hence, detecting B-ALL cells from CNMC_B-ALL is challenging as cancer, and healthy cells appear very similar morphologically.

### 1.3 Summary of Contributions

A summary of contributions from this research study presented below:

1. This study implemented four commonly used features descriptor algorithms to generated a train-test dataset using CNMC_B-ALL dataset for leukemia detection.
2. A shallow CNN with six-layers and around 2M trainable parameters proposed and trained on each feature descriptor individually to compare the model's leukemia detection ability.
3. The proposed S-CNN achieved a classification accuracy of 94.59% on original images and 91.04% on HUE-SIFT images.
4. The experimental results show that proposed S-CNN works well with original pre-processed images for detecting leukemia compared to the feature descriptor used in this study and baseline VGG16 with transfer learning.

## 2. Related Work

### 2.1 Background

The diagnosis of leukemia starts with the analysis of WBC's, mainly based on examining the morphological structure of lymphoblasts under the microscope. Hematologists study blood smear samples to identify positively infected blood cells. Manual detection primarily depends on the skills of the physician. Also, the process of manual discovery can be tiring and time-consuming. Another challenge of the manual detection process is that mature cells appear morphologically similar to immature cells, making the task more challenging. Not all pathology laboratories have skilled hematologists and therefore lack the required resources. Hence, various automated systems have been proposed to identify different blood cells, including immature lymphocytes (Imran Razzak and Naz 2017; Putzu, Caocci, and Di Ruberto 2014).

In the literature, many automated systems have been proposed based on machine and deep learning algorithms with different feature extraction techniques for the classification of the microscopic images. The next section mainly focuses on literature for the classification of ALL microscopic images based on different feature descriptors and classifiers. Most of the past research makes the use of the ALL-IDB dataset for automatic detection of leukemia. This dataset is maintained by M. Tettamanti Research Center for childhood leukemia and hematological illnesses, Monza, Italy (Mishra et al. 2017). Since the dataset CNMC_B-ALL (used in this paper) was released in 2019 by IEEE ISBI, very little research exists on the dataset.

### 2.2 Leukemia detection based on Transfer Learning as feature extraction

The generalization ability of any deep learning model mainly depends on the size of train data used for model training. In general, better results can be achieved on unseen test data if the model is trained with a large number of train data. However, it is very challenging to gather a large amount of data to train and test the model in practice. Hence, deep learning models proposed over the years were trained on the ImageNet dataset (is the collection of a vast number of natural images). These models are known as the pre-trained models.

The main characteristic of these models is after training on natural images best training weights are saved. These saved pre-trained weights then used later on to solve similar problems. Some of the examples of these pre-trained models are VGG16, AlexNet, MobileNet, ResNet, GoogLeNet (Simonyan and Zisserman 2014; Krizhevsky, Sutskever, and Hinton 2012; Howard et al. 2017; He et al. 2016; Szegedy et al. 2015). Therefore, pre-trained models are frequently used for feature extraction based on the concept of transfer learning. The notion of transfer learning is first to train the base model using relevant datasets to perform specific tasks and then transfer the task using the target dataset. The transfer learning process has two steps; the first step is to select a suitable pre-trained model, and the second step is feature extraction from the target dataset. This section presents a literature review for microscopic image classification based on the concept of transfer learning as the feature extraction technique.

The model proposed by Kassani et al. (2019) to classify ALL microscopic images is a hybrid model with transfer learning. They used VGG16 and MobileNet to build a hybrid network. After performing image pre-processing by applying techniques such as resizing, normalization, enhancement, the classification task was performed by the hybrid network. The individual model of hybrid net extracts features from the training

dataset and concatenates them before the fully connected dense layer. This approach's motive was to combine low-level features generated from intermediate layers of the hybrid CNN model to create a high-level discriminative feature map. Using this method, they achieved the classification accuracy of 96.17%.

The solution proposed by Vogado et al. (2018) adopts transfer learning for feature extraction and SVM classifier to identify leukemia. They used three different pre-trained models, namely AlexNet, CaffeNet, and VGG-f, for transfer learning feature extraction. Then feature selection was performed on feature vector generated from transfer leaning to eliminate unnecessary characteristics. The feature selection step uses the grain ratio attribute selection algorithm. The idea behind the grain ratio algorithm is to calculate information gain using entropy to measure the degree of impurity and use a decision tree to represent the attribute vector. Selected features trained on the SVM to achieve an accuracy of 99%.

The transfer learning can be applied in different ways such as use pre-train models to extract features, train top layers of model and freeze out rest of the layers, fine-tune the model parameters of deep learning model or train model from scratch (Vijayalakshmi et al. 2019). The solutions proposed by Goswami et al. (2020) for ALL classification primarily based on a deep learning model and the subject-level heterogeneity loss function. They initialized the weights of Inception-v3 generated by pre-training the model on the ImageNet dataset and used this pre-trained model to extract features. The Inception-v3 has different kernel sizes that help to capture and exploit information on a different scale. Furthermore, the study proposed sampling strategies to capture subject-level heterogeneity and introduced a new loss function based on heterogeneity. The classification stage combines the new loss function with the pre-trained Inception-v3 CNN for training and identification of ALL cells. Using 7-fold cross-validation, they obtained a classification accuracy of 90.05%.

**Table 2**
Literature summary for identification of leukemia using transfer learning as feature extraction

| Study | Transfer Learning Model | Dataset | Accuracy (%) |
|---|---|---|---|
| Kassani et al. (2019) | VGG16 and MobileNet | CNMC_B-ALL | 96.17 |
| Vogado et al. (2018) | AlexNet, CaffeNet and VGG-f | ALL-IDB | 99 |
| Goswami et al. (2020) | Inception-v3 | CNMC_B-ALL | 90.05 |

The transfer learning technique performs well in many cases, but selecting the right pre-trained model for feature extraction can be challenging. Additionally, these pre-trained models trained on natural images; thus, they might not perform that well for every microscopic image datasets. Table 2 summarizes the classification accuracy on two different microscopic image datasets, namely ALL-IDB and CNMC_B-ALL, from previous studies.

### 2.3 Leukemia detection based on Segmentation as feature extraction

The multiple parts or regions of the image can be recognized based on the characteristics of pixels present in an image. The process of separating these regions from a background using digital image processing is called segmentation. Several algorithms and techniques have been proposed and used by researchers over the years for image segmentation. In this section, a brief overview of leukemia detection based on image segmentation as a feature extraction technique from various researches is summarized.

The method proposed by Al-jaboriy et al. (2019) is based on the segmentation of blast cells and the artificial neural network (ANN) to identify positive leukemia blast cells. After segmentation, the system used statistical functions to obtain pixel-based information, and later on, ANN was used to identify positive embryonic cells from negative cells. The accuracy obtained using this method is about 97%.

Alagu and Bhoopathy Bagan (2019) proposed the method of image segmentation based on k-means clustering. Each pixel in the image assigned to a cluster based on the distance between the pixels. The center of each cluster identified by averaging all pixels in the cluster, which leads to cluster-based image segmentation. Then, the authors used GLCM and LBP to extract features such as shape, texture, and color. SVM with Gaussian Radial Basis Function (RBF) produced the best classification results with about 95.30% accuracy.

The leukemia detection method used by Jha and Dutta (2019) was built on segmentation and Chrono-SCA (Sine cosine algorithm). Segmentation techniques such as active counter-based models, fuzzy c-means clustering, and entropy-based hybrid models have been used for image segmentation. The image features, mainly color-based histograms and statistical features such as mean, variance, and standard deviation, were extracted from segmented images to generate the final feature vector. Finally, the Actor-critic neural network and Chrono-SCA algorithm used together to achieve 99% classification accuracy.

Rehman et al. (2018) proposed a solution for classification of ALL cells using deep learning model. The blood images used for research contains different blood components; therefore, to separate lymphoblasts from the rest of the blood cells, they applied segmentation based on a simple threshold method. The implemented CNN consists of five convolutional layers, with two fully connected dense layers. The proposed deep learning model achieved an accuracy of around 97%.

Mishra et al. (2017) proposed an ALL detection method based on the random forest classifier. Before feature extraction, they segmented the images based on the marker-based watershed segmentation method. Next, the grayscale co-occurrence matrix used to extract texture information on the segmented images. The final feature vector was then generated by applying probabilistic principal component analysis (PPCA) and trained on the random forest classifier to achieve up to 96% accuracy on test data.

Rawat et al. (2017a) used a method called hybrid hierarchal classifiers. The CAD system implemented by them consists of several modules such as segmentation module, feature extraction module, and classification module. The first module called segmentation performs pre-processing such as histogram filtering, 2D order-statistic filtering, and segments nucleus based on global thresholding and morphological opening. The feature extraction module extracts feature based on geometric features, chromatic features, and statistical features. Then the final feature vector was generated by applying principle component analysis (PCA). Finally, the classifier module segregates cancer cells from normal cells using different classifiers like SVM, k-NN, probabilistic neural network (PNN). The best classification accuracy of 94% achieved by the SVM classifier.

Table 3 summarizes the various segmentation techniques used for classification of leukemia using ALL-IDB dataset.

**Table 3**
Literature summary for detection of leukemia based on Segmentation using ALL-IDB dataset

| Study | Segmentation Technique | Classifier | Accuracy(%) |
|---|---|---|---|
| Al-jaboriy et al. (2019) | Statistical features with ANN | ANN | 97 |
| Jha and Dutta (2019) | K-mean Clustering | SVM | 95.30 |
| Rehman et al. (2018) | Active counter-based models, Fuzzy c-means clustering, Eentropy-based hybrid models | Actor-critic NN | 99 |
| Mishra et al. (2017) | Threshold method | Random forest | 96 |
| Rawat et al. (2017a) | Global thresholding, Morphological opening | SVM | 94 |

### 2.4 Leukemia detection based on shape, texture and color as features

The information about shape, texture, and color can be extracted from the image using various techniques. Some of the commonly used techniques are histogram equalization, local binary pattern, wavelet coefficients, S-transformation, and Fourier transformation. In general, researchers use these techniques individually or in combination to extract relevant information regarding shape, texture, and color as feature vectors. This section focuses on reviewing leukemia detection based on shape, texture, and color using various techniques.

A method proposed by Mishra, Majhi, and Sa (2019) involves, first pre-processing of an input image to remove background noise by histogram equalization and image thresholding. The final image contains only lymphocytes and lymphoblasts. Second, several features extracted based on morphological structure, texture, and color of images. To extract the morphological features nucleus area, cytoplasm area, lymphocyte area, nucleus perimeter, nucleus shape considered. They used wavelet coefficients, DCT, GLCM (Gray Level Co-occurrence Matrix), GLRLM (Gray Level Run Length Matrix), Fourier transform, and S-transform to extract texture-based features. Third, linear discriminant analysis (LDA) applied to perform feature reduction on texture features. Finally, different classifiers were trained using cross-validation, such as SVM, random forest, K nearest neighbor (KNN), ADBRF. Amongst the different classifiers used, ADBRF performed best, providing about 99% accuracy.

Many novel approaches use shape, color, and texture information to distinguish blast lymphocytes from WBC's. Alongside shape features, Tuba et al. (2019) used a local binary pattern (LBP) algorithm to extract texture information from microscopic images. Finally, the bore bone firework algorithm (BBFWA) used for parameter tuning of the

SVM classifier. In the classification stage, the SVM classifier with the best parameter adjustment used to classify ALL cancer and achieved 91.84% accuracy.

**Table 4**
Literature summary for classification of ALL using different feature descriptor and classifiers (2019-2020)
1 implies ALL-IDB dataset
2 implies CNMC_B-ALL Dataset

| Research | Model | Feature Descriptor | Dataset | Accuracy |
|---|---|---|---|---|
| Kassani et al. (2019) | Hybrid Net with VGG16 MobileNet | Transfer Learning | 2 | 96.17% |
| Al-jaboriy et al. (2019) | ANN | Segmentation, Statistical features | 1 | 97% |
| Mishra, Majhi, and Sa (2019) | ADBRF | Shape, Texture, Histogram normalization | 1 | 99% |
| Tuba et al. (2019) | SVM | LBP | 1 | 91% |
| Alagu and Bhoopathy Bagan (2019) | SVM | Cluster Based image Segmentation | 1 | 95% |
| Goswami et al. (2020) | Inception-v3 | Transfer Learning with Heterogeneity Loss Function | 2 | 95.24% |
| Ahmed et al. (2019) | CNN | Data Augmentation | 1 | 88% |
| Jha and Dutta (2019) | Actor Critic NN | Segmentation | 1 | 99% |
| Jung et al. (2019) | W-NET | Preprocessed Images | 1 | 96% |

### 2.5 Leukemia detection using simple deep learning models

It is possible to achieve great classification results with a simple model if trained with a large number of images. Many scientists use the data augmentation technique to create a synthetic dataset. This section summarizes a simple neural network with the data augmentation technique for the detection of cancer cells.

The data augmentation technique is a commonly used approach to increase the dataset and overcome the deep learning model's memorization problem. Moreover, this approach reduces the error rate to obtain a more generalized classification model. The methods proposed by Ahmed et al. (2019) use data augmentation techniques such as rotation (40°), height offset (40%), width offset (40%), zoom (30%), horizontal flip, vertical flip and shearing (20°); hence the resulted dataset increased by seven times. A CNN model with two convolutional layers, two max-pooling layers, and a fully connected dense layer implemented and trained with various optimizers. A model with the *Adam* optimizer produced classification accuracy of 88% on train data.

The w-net proposed by Jung et al. (2019) for detection of WBC's has a simple architecture with three convolutional layers followed by two fully connected layers. The authors applied standard pre-processing techniques such as resizing, cropping, and normalization on the dataset. In the last step, w-net uses 10-fold cross-validation to obtain about 96% accuracy. Thanh et al. (2018) used CNN based classification approach for differentiating ALL cells from normal cells. They used the image augmentation and CNN to achieve a classification accuracy of 97%.

Table 4 represents the literature summary for the classification of leukemia from the year 2019 and 2020. The table describes the various feature extraction techniques and models used for leukemia detection. The two main datasets used by researchers for their studies are the ALL-IDB dataset and CNMC_B-ALL dataset.

## 3. Experimental Setup

This section explains the various experimental setup steps required to identify leukemia blood cells from the microscopic images. Figure 1 explains the logical flow of the experimental setup. The first section describes the dataset used in this study, followed by the explanation of the image feature descriptors. The middle section provides a detailed explanation of the architecture of the proposed S-CNN. The final section reveals the sampling method, evaluation metrics, evaluation baseline, hardware setup, and model training.

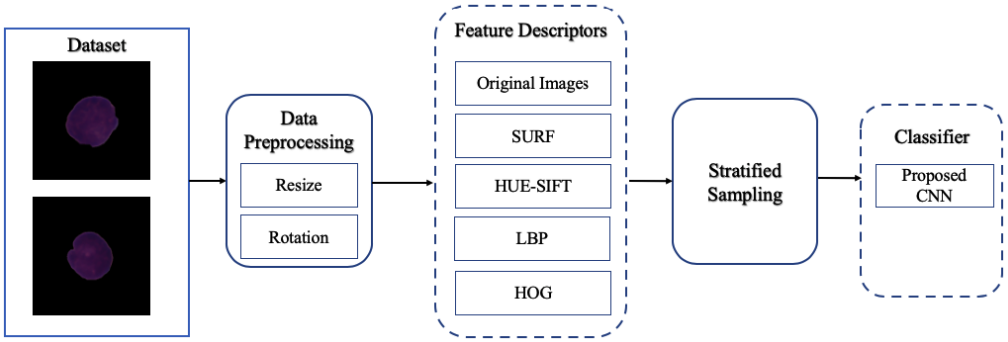
### 3.1 Dataset Description

In this paper, the dataset from a challenge held by IEEE ISBI called *Classification of Normal versus Malignant Cells in B-ALL White Blood Cancer Microscopic Images* is used. The dataset is available public and maintained by SBILab. It consists of microscopic blood images segmented from WBC's of 69 ALL patients and 49 healthy subjects. Each image consists of single-cell cancer or healthy WBC separated from blood smear images that are captured from the bone marrow sample slides. As per the description of the dataset from the readme, the staining color variations of the images are standardized, and the cells labeled by professional oncologists. In total initial dataset consists of 7272 cancer and 3389 normal blood cell images (SBI 2018).

**3.1.1 Data pre-processing: Resizing.** The single image size of the original dataset provided by the SBILab is 450 * 450 RBG pixels. The original images were resized to 150 * 150 pixels without compromising image quality for a smaller representation feature vector.

**3.1.2 Data pre-processing: Rotation.** The lack of sufficient data for the classification task is a common problem especially true in the case of medical imaging. Thus, many re-

**Figure 1**
Logical flow of experimental Setup




searchers use a novel approach of image augmentation to augment, generate, or balance data sets. A widespread practice for image augmentation is to perform geometric or color change operations such a horizontal flip, vertical flip, cropping, change in a color pallet of the image (Frid-Adar et al. 2018). In the CNMC_B-ALL dataset compared to the cancer patient images, healthy cell images are low in number; therefore, the dataset is highly imbalanced. Thus, to overcome this problem, the non-cancer images rotated to 180 degrees to generate more healthy cell images.

### 3.2 Image Feature Descriptors

This section briefly explains the workflow of each feature descriptors used in this study to generate new feature based images.

**3.2.1 Speeded Up Robust Features:.** In 2006, Bay, Tuytelaars, and Van Gool introduced speeded up robust features (SURF) algorithm based on the invariant feature transform algorithm that was first suggested by Lowe in 2004. The extraction of SURF features from the image involves multiple steps. The initial stage of SURF generates keypoints. Next, based on these generated keypoints, orientation invariant descriptors determined.

The objects in the image differentiated by the abrupt change in the direction of the boundary's edge. This sudden change considered as one or more points of interest. The intersection between two or more edge line segments also regarded as a keypoint. SURF approximates the keypoints in the image based on the Hessian matrix. This approximation is scale and rotation invariant (Srivastava, Bakthula, and Agarwal 2018). To generate the SURF features, the area around the detected keypoints first divided into multiple sub-areas, and then the wavelet response calculated for each sub-area and represented as a feature vector (Karami, Prasad, and Shehata 2017).

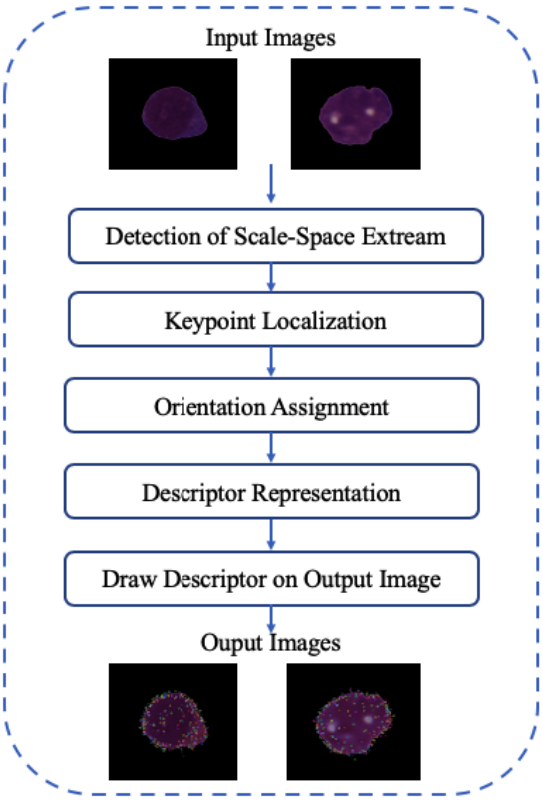
Suppose $I$ is the region around a point $P(x, y)$, then Hessian matrix $H(P, s)$ at the scale $s$ defined as follows:

$$\begin{bmatrix} Z_{xx}(P,s) & Z_{xy}(P,s) \\ Z_{yx}(P,s) & Z_{yy}(P,s) \end{bmatrix} \quad (1)$$

where $Z_{xx}(P,s), Z_{xy}(P,s), Z_{yx}(P,s), Z_{yy}(P,s)$ are the convolution of the Gaussian second order derivative of the image $I$ at a point $P$. For each pixel in the image determinant of Hessian calculated and using those determinant values, keypoints are determined (Srivastava, Bakthula, and Agarwal 2018).

In this study, the same method is used to calculate the keypoints of microscopic images, and then computed keypoints were drawn on original images, as shown in Figure 2.

**Figure 2**
Process flow to generated SURF images from pre-processed images based on SURF algorithm




**3.2.2 Scale Invariant Feature Transform.** One of the robust and highly efficient feature descriptor algorithm for the image with respect to rotation, scale change, affine transform, and viewpoint change is scale-invariant feature transform (SIFT), which was first introduced in 2004 by Lowe (Karami, Prasad, and Shehata 2017). Before extracting features from microscopic images, the hue of the image is adjusted for best color pixel representation, and then the SIFT algorithm applied. The four-step SIFT feature extraction process represented in Figure 3 is explained briefly in the below sections:

**Step1: Scale-space extrema detection:** In the first step, different viewpoints of the image, such as rotation, translation, and scaling, are used to identify keypoints. These keypoints are invariant to the scale. The main operation in this step is to apply a Gaussian filter that calculates the difference of Gaussian (DoG) to obtain the scale space.

The Gaussian filter function $G(x, y, \sigma)$ applied to the image $I(x, y)$ results in the scale-space $L(x, y, \sigma)$, see equation 2:

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y) \tag{2}$$

where x, y are coordinates of the input image *I* for a given pixel, and $\sigma$ is the width of Gaussian filter *G*. Once scale-space is calculated, the DoG of the image is obtained using function 3:

$$D(x, y, \sigma) = L\left(x, y, k_i, \sigma\right) - L\left(x, y, k_j, \sigma\right) \tag{3}$$

where $k$ is the multiplicative factor from two different viewpoints of the same image. Finally, local maxima or minima from the DoG images across scale represented as keypoints on the image (Gheisari et al. 2018; Patil and Sinha 2016).

**Step 2: Keypoint Localization:** Not all of the extracted keypoints represent relevant features as they suffer from low contrast and might have poorly localized edges. Such keypoints are removed by calculating Laplacian. A contrast threshold $\mathrm{C_C}$ is set, and if the value of Laplacian function is less than $\mathrm{C_C}$ those points are rejected. Therefore, after eliminating such keypoints, more stable low contrast edge points are gathered (Gheisari et al. 2018; Patil and Sinha 2016).

**Step 3: Orientation Assignment:** In this step, keypoints that are extracted in the previous step assigned to orientation to make them invariant to rotation. To achieve this, the orientation of the histogram of local gradient is applied to Gaussian smoothed image $L(x, y, \sigma)$. First, it uses the pixel difference to calculate the gradient magnitude $m(x, y)$ and direction $theta(x, y)$ for the region around the keypoint, as shown below:

$$\theta(\mathrm{x}, \mathrm{y}) = \tan^{-1}((L(x, y+1) - L(x, y-1)/(L(x+1, y) - L(x-1, y)) \tag{4}$$

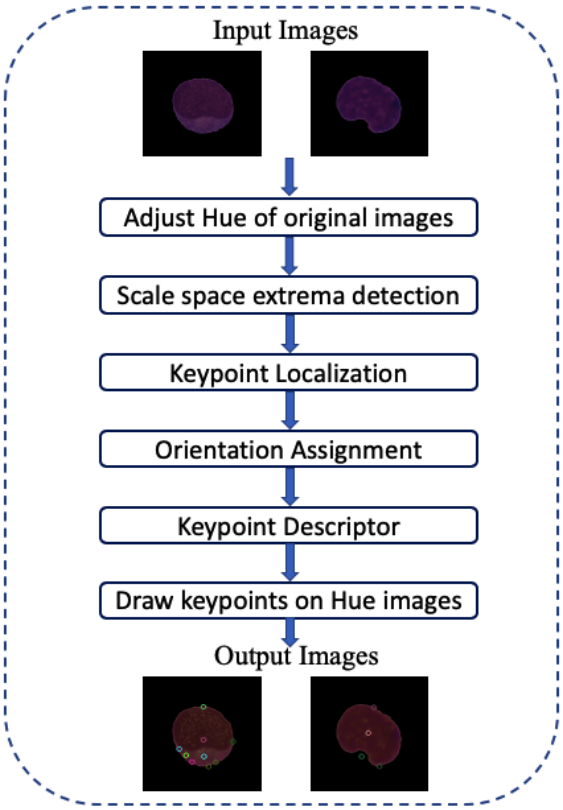
$$m(x, y) = \sqrt{(L(x, y+1) - L(x, y-1))^2 + (L(x+1, y) - L(x-1, y))^2} \tag{5}$$

Then, an orientation histogram with 36 bins is formed and converted into a 360-degree orientation. These bins consist of samples from neighboring windows of keypoint, which are weighted by the gradient magnitude $m(x, y)$ and circular gaussian with $\sigma$ equal to 1.5 times the scale of the keypoint. Histogram peaks higher than 80% are regarded as the highest peaks and assigned to the keypoints (Gheisari et al. 2018; Patil and Sinha 2016).

**Step 4: Keypoint Descriptor:** In the last step, a SIFT descriptor is calculated for each keypoint as a histogram of orientation. Typically, the region around the of size 16 ×16 divided into 4×4 smaller regions and aggregated together to form an 8-bin histogram. Then, SIFT feature vector of size 4 × 4 × 8= 128 is generated from each keypoint (Gheisari et al. 2018; Patil and Sinha 2016).

Before running the SIFT algorithm, three parameters are required to set in advance,the width of the Gaussian $\sigma$, the contrast threshold $\mathrm{C_C}$ and the edge threshold $\mathrm{C_E}$. The final HUE-SIFT images are generated by adjusting the hue of original images, then

**Figure 3**
Steps to generate HUE_SIFT images based on SIFT algorithm




from these hue shifted images keypoints are computed by applying the SIFT algorithm and drawn on hue images. These HUE-SIFT images are used to train the S-CNN.

**3.2.3 Local Binary Pattern.** Another well-known feature description technique mainly used for pattern recognition and classification tasks,is the Local Binary Pattern (LBP). Ojala, Pietikäinen, and Harwood first introduced LBP in 1996. LBP helps to extract texture information from images very efficiently. The characteristics of LBP are **a)** robust against illumination changes, **b)** very fast to compute, **c)** extracts local features, and **d)** very few initial parameter settings required, which makes it extensively used feature descriptor. The main principle behind LBP is to detect local patterns between adjoining pixels (Naufa and Sajith 2020).

The detailed process used to extract LBP features in this study explained in Figure 4. Given two parameters, radius $R$ and number of pixels $P$, center $Q_c$ is considered. Then, the distance between the center $Q_c$ and $P$ number of neighbours $(Q_l, Q_2, Q_3, \ldots\ldots, Q_{P-1})$ is calculated and converted into binary number. Later, the generated LBP code is assigned to these neighbours using following two equations:

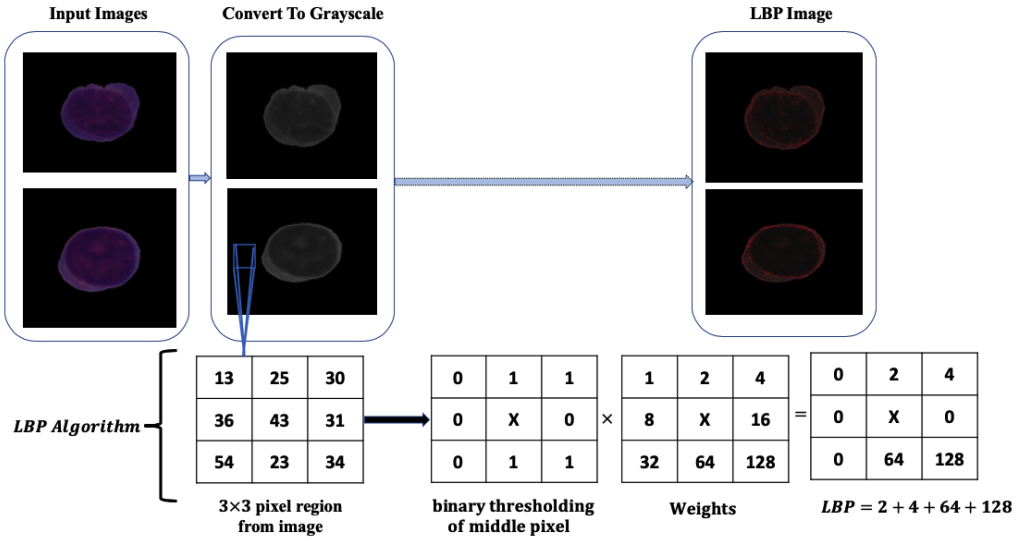
$$f(x) = \left\{ \begin{matrix} 1, x \geq 0 \\ 0, \quad x < 0 \end{matrix} \right\} \tag{6}$$

$$LBP_{P,R} = \sum_{P=0}^{P-1} f\left(Q_P - Q_C\right) 2^P \tag{7}$$

At the given threshold (in this case, 0), equation 6 produces one if the difference between two points is greater than the threshold, otherwise it produces zero (Liu et al. 2017).

In this study, images generated using LBP features used to train a proposed neural network. After calculating LBP codes, they are drawn on a grayscale image to produce LBP images.

**Figure 4**
LBP image generation flow using LBP algorithm




**3.2.4 Histogram of Oriented Gradient.** This section briefly explains the working of the Histogram of Oriented Gradient (HOG) algorithm used for generating HOG images. This method was first introduced by Dalal and Triggs for human detection from video in 2005. The principle behind HOG is that the appearance and shape of objects in an image can be represented by the intensity gradient and edge orientation distribution. The four steps involved in the HOG algorithm explained in this section.

According to Adetiba and Olugbara (2015), the first step calculates gradient values by applying a one-dimensional derivative mask or the finite difference approximation. The input image is filtered using one-dimensional derivative in both vertical and horizontal direction with the kernel in

$$D_x = [-1, 0, 1] \tag{8}$$

$$D_y = [1, 0, -1] \tag{9}$$

Then, on grayscale image convolution operation is performed to obtained x and y derivatives as follows:

$$I_x = I \times D_x \tag{10}$$

$$I_y = I \times D_y \tag{11}$$

where $I$ is a input grayscale image. Also, magnitude and orientation of the grayscale image are computed respectively as below

$$|G| = \sqrt{I_x^2 + I_y^2} \tag{12}$$

$$\theta = \tan^{-1}\left(\frac{I_x}{I_y}\right) \tag{13}$$

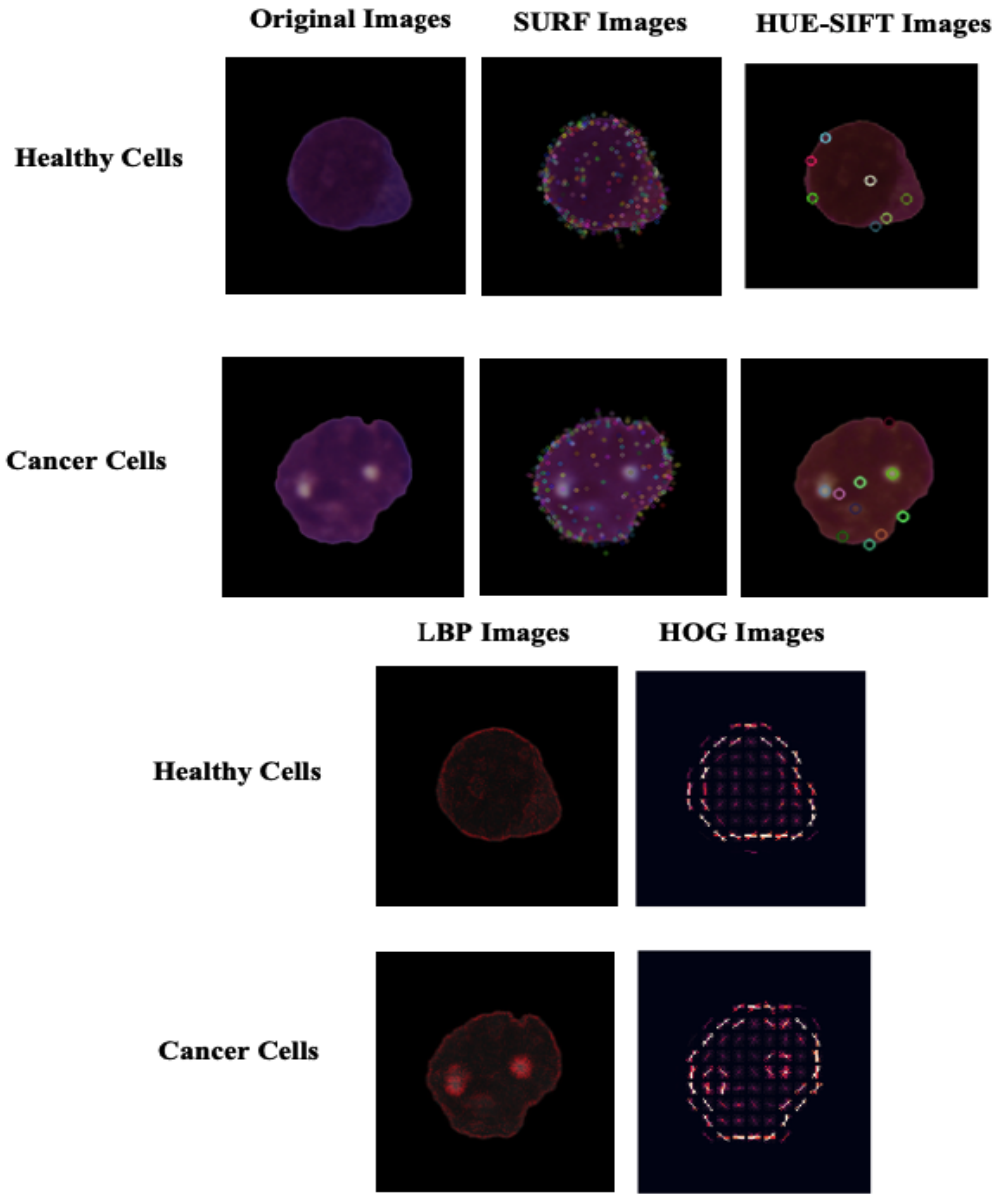
In the second step (orientation binning), HOG creates the cell histograms. First, it uses a rectangular HOG cell and calculates the gradient values. Then, based on the histogram channel, the weight is assigned to each pixel present inside the rectangular cell. Histogram channels can be signed or unsigned. Signed channels mainly distributed between 0 to 180 degrees and unsigned channels mainly distributed between 0 to 360 degrees (Adetiba and Olugbara 2015).

The descriptor block is created in the third step of the HOG algorithm, and then it is combined with all orientation bin generated in the second step to create larger spatially connected blocks. Since the collected bin forms the larger blocks, it accounts for changes in illumination and contrast. In the last step block normalization is applied to these blocks such as *L1* or *L2* normalization (Adetiba and Olugbara 2015). For this study, *L2* normalization is used for block normalization, as shown below:

$$L2_{\text{norm}} f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{14}$$

Figure 5, summarizes the images generated using different feature descriptors which are used to train the proposed S-CNN.

**Figure 5**
Representation of images generated from each feature descriptor. First row of each column represents healthy lymphocyte images and second row of each column represents cancer lymphocyte images




### 3.3 Architecture of the proposed shallow Convolutional Neural Network

The basic building blocks of the proposed S-CNN are convolutional layers, pooling layers, normalization layers, dropout layers, and fully connected dense layers. The

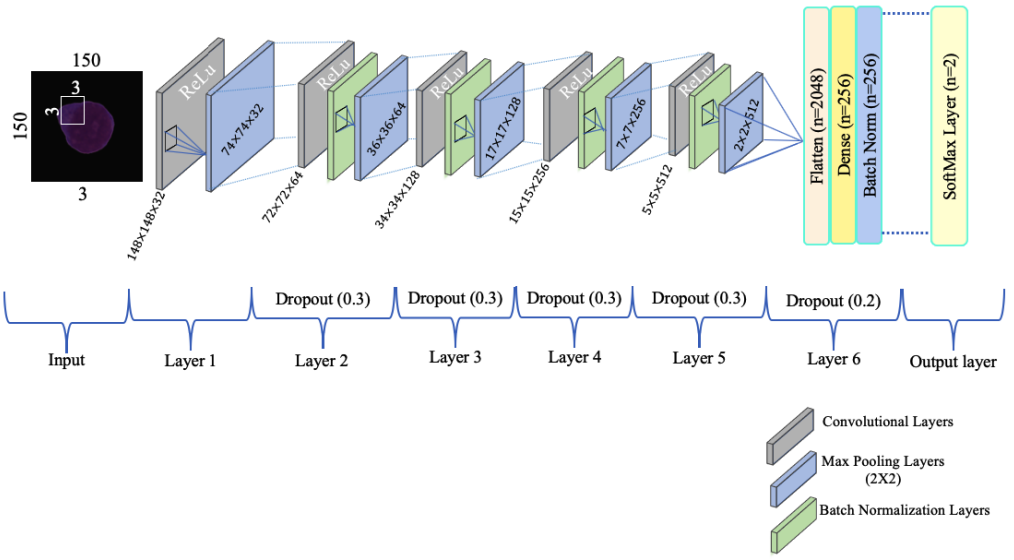
schematic representation of the proposed S-CNN presented in Figure 6 and explained in detail below:

**3.3.1 Convolutional Layers.** The S-CNN used in this study consists of five convolutional layers. The first convolution layer has 32 filters; the second layer has 64 filters; the third layer has 128 filters; the fourth and fifth layer have 256 and 512 filters. Each convolutional layer has a 3×3 kernel filter and uses a rectifier linear unit (*ReLu*) as an activation function. Given the input, the convolutional layer generates output neurons by calculating the weighted sum of input neurons, adding bias to that weighted sum, and apply the activation function to it. From different variants of *ReLu* activation function, $Relu(x) = \max(0, x)$) is used with the proposed S-CNN. The main function of each convolutional layer is to extract and learn features from input images.

**3.3.2 Max pooling and Batch Normalization layers.** Each convolutional layer followed by a batch normalization layer and a max-pooling layer. All max-pooling layers have a 2×2 filter. The batch normalization layer uses the mean and standard deviation of the current mini-batch to normalize the node's activation. With decreasing dimensions from filtered images, the max-pooling layer focuses on the image's essential features or objects.

**3.3.3 Dropout, Flatten and Fully connected Dense Layers.** The over-fitting was avoided by adding the dropout layers after max-pooling and fully connected dense layers, and the dropout rate set to $p = 0.3$ and $p = 0.2$, respectively. The output generated after the last block of convolution-normalization-pooling layers, it is flattened to a one-dimensional feature vector using the flatten layer. The output of the flatten layer with 2048 neurons is used as input for the first fully connected dense layer. The first dense layer has total 256 neurons with *ReLu* activation function. The second dense layer is used for classification with *softmax* activation function.

**Figure 6**
Six layer architecture of proposed convolutional neural network (S-CNN)



### 3.4 Sampling

The original dataset is first divided into train (90%) and test (10%). Then a stratified sampling technique with 7-fold cross-validation is used to train the model. During cross-validation, the train dataset further divided into train and validation subsets. The random stratified sampling method ensures that each training fold has a diverse train-validation dataset.

### 3.5 Evaluation Metrics:

The efficiency of the proposed methodology evaluated using four medically significant statistical measures, namely overall accuracy, sensitivity, specificity, and weighted F1-score. To calculate these metrics, the confusion matrix generated based on the table presented in Appendix 5.5. In this study, positive is cancer cells, and negative is healthy cells. So true positive (TP), false positive (FP), false negative (FN), true negative (TN) are defined as follows:

- TN: correctly identified noncancer cells
- FN: incorrectly identified noncancer cells
- FP: incorrectly identified cancer cells
- TP: correctly identified cancer cells

**Accuracy:** The overall classification accuracy is obtained by adding TP and TN, and by dividing it with total samples.

$$\text{Accuracy } = \frac{TP + TN}{TP + FP + FN + TN} \tag{15}$$

**Sensitivity:** The measure of proportion of actually identified cancer cell images out of total cancer cell images.

$$\text{Sensitivity or Recall } = \frac{TP}{TP + FN} \tag{16}$$

**Specificity:** The measure of proportion of actually identified healthy cell images out of total healthy cell images.

$$\text{Specificity } = \frac{TN}{TN + FP} \tag{17}$$

**Weighted F1-Score:** It is the calculation of the F1-score of each class and adding them together based on the weight of true labels of each class. F1-score is calculated using precision and recall given below.

$$\text{precision} = \frac{TP}{TP + FP} \tag{18}$$

$$\text{F1 score} = 2\frac{\text{precision} * \text{recall}}{\text{recall} + \text{precision}} \tag{19}$$

### 3.6 Evaluation Baseline

The experimental results were compared with the novel approach of transfer learning by creating a baseline using the VGG16 model. First, using a pre-trained model VGG16 transfer learning feature vector of size 512 neurons is generated. All trainable layers of original VGG16 is frozen and two fully connected dense layers added, the first dense layer has 512 neurons, *ReLu* activation function, and the input shape of 512. The second dense layer has two classes with activation function *softmax*. Results generated from this baseline setup is used to compare against the results of the experimental setup of this study. The motivation behind this approach is that previous research mainly uses the ALL-IDB dataset for ALL classification using transfer learning and VGG16. The single image from this dataset consists of multiple numbers of blood cells; therefore, immature cells are readily detectable from clusters. However, the dataset used in this study consists of a single-cell per image, making it challenging to differentiate cancer cells from healthy cells since they have the same morphological structure. Therefore, the baseline and the experimental results have same underlying dataset for valid comparison.

### 3.7 Hardware and Model Training

The designed experiment is implemented in Python using Keras package (Chollet et al. 2015) to generate proposed S-CNN. The CNMC_B-ALL dataset divided into the train (90%) and test (10%). Then, the proposed model trained on a training dataset using stratified cross-validation with 7-folds. The hardware used for training is Tesla K80 GPU with 12GB memory 61GB RAM and 100GB SSD hosted on the cloud. The total training time of S-CNN per feature descriptor is around 90 minutes. The python code used to generate experimental results in this study is available at Appendix 5.5

The model was compiled and trained using different optimizers such as adaptive movement estimation *(Adam)*, stochastic gradient descent *(SGD)* with momentum, and root mean square propagation *(RMSProp)*. Furthermore, this study experimented with two different loss functions, namely binary cross-entropy and sparse categorical cross-entropy, to calculate the error rate. The setup with sparse categorical cross-entropy loss function and *Adam* optimizer produced the best results on training data. The final predictions were obtained using the test dataset and compared against the baseline. Training weights from each 7-folds saved, and final results on test data generated from best fold weights.

## 4. Results

In the comparative analysis of feature descriptors for leukemia detection, this study performed five experiments. Each experiment follow the individual training of S-CNN on five different sets of train images. The train images were generated by applying various feature descriptor algorithms on the CNMC_B-ALL dataset. A stratified cross-validation sampling technique was applied to train S-CNN. The best training results were obtained by parameter tuning with different values of k (number of folds) in cross-validation. Out of 5,7,10 parameter values of k, the model produced the best results

with k=7. The final prediction results achieved from unseen test data of the CNMC_B-ALL dataset. The test data consists of 1406 images with 728 cancer images and 678 healthy images. This section describes experimental results acquired from each feature descriptor trained with S-CNN individually.

Table 5 represent the validation accuracy for each feature descriptor achieved by S-CNN. The first row in the Table 5 represents the best accuracy results on validation data out of 7 folds. The second row summarizes the average accuracy of 7-fold cross-validation with standard deviation. Figure 7 summarizes the total number of positively and negatively predicted cancer and healthy cells from microscopic images. The first bar graph reveals the number of correctly predicted cancer cells (blue) and incorrectly predicted cancer cells (orange). Similarly, the second bar graph shows the number of positively identified healthy cells (blue) and negatively identified healthy cells (orange).

### 4.1 Experiment 1: S-CNN trained on RAW Images

In the first experiment, the proposed S-CNN was trained on pre-processed original images. From Table 5, the best accuracy achieved on validation data is 99.17%. The generalization capability of the model tested based on the predictions generated using test data with 1406 images. The proposed system reported 94.59% accuracy on test data. Furthermore, this experiment achieved 95.33% of the sensitivity rate and 93.81% of the specificity rate.

### 4.2 Experiment 2: S-CNN trained on SURF Images

In the second experimental setup, classification results obtained using images generated from the SURF feature descriptor. For SURF images, the keypoints were detected using the SURF algorithm and drawn on the original image to obtain the final SURF image dataset. As presented in Table 5 the proposed S-CNN with SURF images achieved the best accuracy of 98.78% on validation data. However, the system achieved only 86.56% of accuracy on test data. The sensitivity and specificity rates reported by the model are 93.82% and 78.76%, respectively.

**Table 5**
Validation Accuracy for each feature descriptor

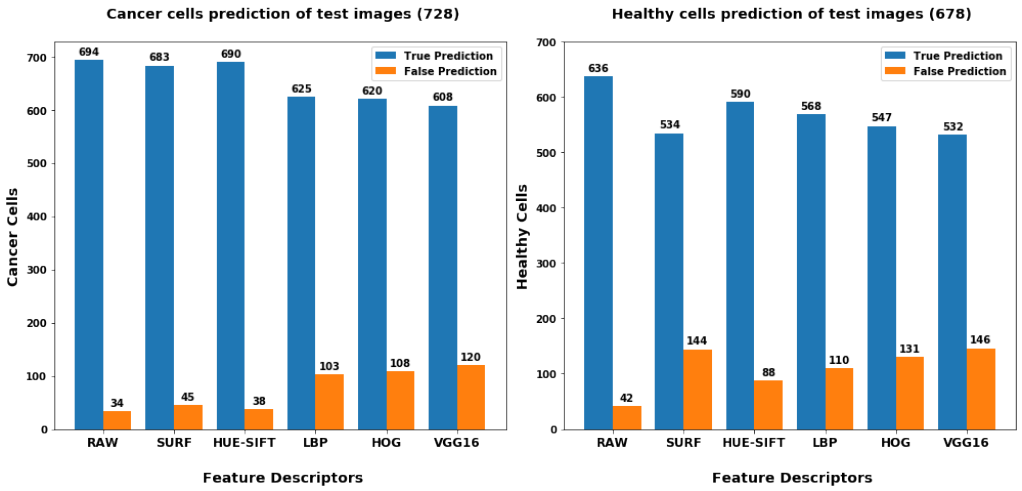
| Feature Descriptors | RAW | SURF | HUE-SIFT | LBP | HOG |
|---|---|---|---|---|---|
| Best Accuracy (%) | 99.17 | 98.78 | 99.78 | 99.39 | 99.56 |
| Average accuracy (%) | 92.72 (+/- 5.01) | 89.66 (+/- 7.36) | 92.29 (+/- 9.19) | 88.80 (+/- 9.85) | 85.41 (+/- 10.63) |

### 4.3 Experiment 3: S-CNN trained on HUE-SHIF images:

In the third experiment, the proposed classifier was trained with HUE-SIFT images. The images can be represented using different color spaces. In this setup, the hue of the original images was adjusted for the best color space representation of the blood smear images. Later, to the HUE adjusted images SIFT algorithm was applied to calculate keypoints. Finally, on these images, computed keypoints were drawn to generate input

images. For this experiment classification accuracy of 99.78% (Table 5) was achieved on validation data. Test data revealed 91.04% accuracy, 94.78% of sensitivity rate, and 87.02% of specificity rate.

**Figure 7**
Predictions on the test data using S-CNN. The graph on the left shows the number of correctly and incorrectly identified cancer cells and the graph on the left shows the number of correctly and incorrectly identified healthy cells.




### 4.4 Experiment 4: S-CNN trained on LBP images

In the fourth experimental setup, S-CNN was trained on LBP images generated from the LBP algorithm using original images. Even though the proposed model performed well on validation data with 99.39% accuracy, it performed poorly on test data with only 84.85% of accuracy. The sensitivity and specificity rates recorded for this experimental setup are 85.85% and 83.78%, respectively.

### 4.5 Experiment 5: CNN trained on HOG Images

In the final experimental setup of this study, S-CNN was trained on HOG features descriptor images. The HOG images are grayscale images generated from original images using the HOG algorithm. The best accuracy of 99.56% obtained on validation data, but the model performs poorly on test data with only 83% accuracy. The sensitivity achieved for the final setup is 85.16%, and specificity is 80.68%.

## 5. Discussion

Table 6 summarizes the results obtained for each experiment based on the evolution metrics used in this study. The first column of the table represents the various feature descriptor used to train S-CNN for the detection of a cancer cell from microscopic images. Due to clinical significance, the analytical study was conducted, taking into account sensitivity and specificity rates along with classification accuracy. This chapter

**Table 6**
Summery of classification metrics on Test Data

| Feature Descriptor | Test Data Metrics (%) | | | |
|---|---|---|---|---|
| | Accuracy | Sensitivity | Specificity | Weighted *F1*-score |
| Original Images | 94.59 | 95.33 | 93.81 | 94.59 |
| SURF Images | 86.56 | 93.82 | 78.76 | 86.46 |
| HUE-SIFT Images | 91.04 | 94.78 | 87.02 | 91.02 |
| LBP images | 84.85 | 85.85 | 83.78 | 84.85 |
| HOG Images | 83 | 85.16 | 80.68 | 82.99 |
| VGG16 with transfer learning | 81.08 | 83.52 | 78.47 | 82.05 |

represents the detailed analysis of each experiment to conduct a comparative study of feature descriptors with S-CNN. Later on, this section compares the performance between S-CNN with feature descriptor and VGG16 with transfer learning as a feature extraction technique.

### 5.1 Answering the Research question and sub-questions

From the experimental results, it can be concluded that S-CNN performs best on pre-processed original images compared to images processed with SURF, HUE-SIFT, LBP, and HOG feature descriptors. Therefore, to answer the research question of this study, even though feature descriptors have good classification accuracy (between 83 to 91%), S-CNN performs best on original images with the highest accuracy of 94.59%. The sensitivity and specificity rates achieved with original images are 95.33% and 93.81%, respectively. During the experiment, it is evident that images generated from different feature descriptors have a loss of important feature information, thus affecting the prediction ability of S-CNN.

The feature descriptor SURF generates the images by drawing keypoints on the original images using the SURF algorithm. From the SURF images in Figure 5, it can be seen that the keypoints add noise to images, with the consequence of losing vital information that helps to distinguish a normal cell from a cancer cell. The S-CNN efficiently identifies cancer images using SURF images with a sensitivity rate of 93.82%, but it performs poorly to identify healthy cells, and the specificity rate is only 78.76%. Therefore, the results from the first experiment suggest that the SURF algorithm is not the right choice for leukemia detection because it adds considerable noise to images resulting in loss of information regarding shape and texture. Due to this, the classifier fails to identify healthy images efficiently.

The hue technique applied to images in the second experiment helped with better presentation of color space in an image. The HUE-SIFT images (Figure 5) are generated by adjusting hue and applying the SIFT algorithm. The HUE-SIFT images performed well on S-CNN with an overall classification accuracy of 91.04% and sensitivity rate of 94.78%. The experimental results suggest that adjusting the hue of images helped somewhat in the classification task, but keypoints drawn on the image using the SIFT algorithm did not play a significant role. Even though keypoints help to represent

localized edges in an image, they do not distinguish healthy cells efficiently compare to original images.

The third experiment generates LBP images (Figure 5) using the LBP algorithm. The idea behind the LBP algorithm is to extract texture information from images using local pixel information. It has been observed that presenting microscopic blood cell images just using texture information does not perform well on S-CNN for leukemia detection. The generated LBP images effectively describe the shape of the blood cell present in an image, but it loses information about cytoplasm present in the cell. When an image is presented on grayscale, it does not have precise information about cytoplasm, which separates the nucleolus inside the cell from outer shape. Hence, S-CNN performs very poorly on LBP images with sensitivity and specificity rates of 85.85% and 83.78%, respectively.

Out of all feature descriptors used in this study, images generated from the HOG algorithm performed the worst on S-CNN. The HOG is considered one of the most effective algorithms for representing local object appearance and shape based on the distribution of intensities or edge direction. The HOG images in Figure 5 represent the outer shape of the blood cells efficiently, but similar to LBP images, they do not represent the inner nucleolus effectively, thus resulting in poor performance on the classifier. The sensitivity and specificity rate obtained in the last experiment using HOG images are 85.16% and 80.68%, respectively.

Another goal of this study was to implement a neural network that will be lighter to train compare to pre-trained models. To achieve this goal, the study successfully implemented a six-layer S-CNN model with 2,096,066 trainable parameters and 2,432 non-trainable parameters. The trainable parameters of proposed S-CNN are less contrast to pre-trained models such as AlexNet (60M train parameters) and VGG16 (130M train parameters). Compared with VGG16's 528MB size, the training weight size of the proposed model is only about 24MB. The training time required by the S-CNN ranges from 60 minutes to 90 minutes, depending on input size and feature descriptor used.

### 5.2 Comparative analysis with baseline and literature

From the literature, most of the studies explore transfer learning with pre-trained models using the ALL-IDB dataset to identify cancer cells. Compared with the ALL-IDB image, the data set used in this study has a single cancer cell per image, while the ALL-IDB image aggregates immature cancer cells in one image. Hence, to perform a comparative analysis between the proposed system and pre-trained models with transfer learning, the baseline was created on the CNMC_B-ALL dataset. From the results, the proposed model performed very well compared to transfer learning with VGG16. From Table 6, the best accuracy achieved by the proposed system is 94.59% on raw data and 91.04% on HUE-SIFT data. In contrast to the baseline accuracy of 81.08%, the proposed S-CNN performed very well. Moreover, all feature descriptors with the S-CNN produced better results compared to the baseline.

Kassani et al. (2019) proposed a hybrid model that uses VGG16 and MobileNet to classify leukemia using the CNMC_B-ALL dataset. They achieved 96.17% accuracy and 95.17% sensitivity. Compared to the classification accuracy of 94.59% obtained by S-CNN, the hybrid model performs better. Therefore, to conclude, feature extraction using multiple pre-trained models is a good approach for leukemia detection. However, the S-CNN identifies cancer cells better than the hybrid model with a sensitivity of 95.33%. Tuba et al. (2019) proposed a method that uses LBP features with SVM classifier. With LBP features and the SVM classifier, they achieved 91.84% of classification accuracy.

However, in this study, LBP images with S-CNN achieved only 84.85% classification accuracy. Hence, LBP features work better with the SVM classifier than S-CNN.

### 5.3 Limitations

This section highlights some of the limitations that occurred while conducting this study. Feature concatenation is the technique of combining various features extracted using feature descriptors. Due to the restriction of hardware resources, this study could not explore this technique. The hardware used in this study lack the capacity to train S-CNN on the dataset that was generated after combine images from each feature descriptor. The various feature descriptor algorithms used in this study uses multiple parameters to create final images. Due to time constraints, the parameters on each feature descriptor were not tuned, and the default parameter values were used. For example, the SIFT or SURF algorithms have a parameter that decides the number of keypoints drawn on images. In this study, these algorithms used default values to draw keypoint and could not experiment with various values of keypoints. Another limitation was that the dataset used in this study is an imbalanced dataset with fewer healthy cells than cancer cells. The dataset was balanced by applying the rotation method of the data augmentation technique. But other data augmentation methods such as vertical flip, horizontal flip, shearing, cropping was not used due to hardware constraint.

### 5.4 Future Work

The performance of the deep learning model can be improved if trained on a larger dataset. The S-CNN proposed in this paper trained using only around 12,000 images; however, its prediction ability can be enhanced if trained on more blood smear images. Data Augmentation is one approach that can help to create a synthetic dataset using CNMC_B-ALL dataset. The data augmentation methods mentioned in the limitation section can be used to generate a larger dataset. In future work, S-CNN can be trained on the datasets generated from the data augmentation for better detection of leukemia. In this research, four different feature extraction techniques were used to classify leukemia; however, instead of using them individually to train the S-CNN feature concatenation approach can be used if better hardware is available. Moreover, the proposed S-CNN can be trained on other microscopic image datasets available publicly to generate pre-trained weights specific to microscopic images.

### 5.5 Conclusion

To summarize, this research paper aimed to conduct a comparative study of feature descriptors for the automatic detection of leukemia cells from microscopic images. Alongside this, the study proposed a shallow convolutional neural network (S-CNN). The experimental results revealed that proposed S-CNN trained on original images produces the best classification accuracy of 94.59%. The HUE-SIFT images generated using the SIFT algorithm performed best out of all feature descriptors used in this study with an accuracy of 91.04%. The results demonstrate that S-CNN identifies cancer cells efficiently from microscopic images compared to VGG16 with transfer learning. Compare to the state-of-the-art pre-trained models, S-CNN is lighter in terms of architecture hence requires less time to train and simple hardware setup. In future work, S-CNN can be further trained with images obtained using techniques such as feature concatenation, data augmentation to improve the performance of the model.

## References

2018. Classification of normal vs malignant cells in b-all white blood cancer microscopic image: Isbi 2019.

Adetiba, Emmanuel and Oludayo O. Olugbara. 2015. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. *The Scientific World Journal*, 2015:1–17.

Ahmed, Nizar, Altug Yigit, Zerrin Isik, and Adil Alpkocak. 2019. Identification of leukemia subtypes from microscopic images using convolutional neural network. *Diagnostics*, 9:104.

Al-jaboriy, Saif S., Nilam Nur Amir Sjarif, Suriayati Chuprat, and Wafaa Mustafa Abduallah. 2019. Acute lymphoblastic leukemia segmentation using local pixel information. *Pattern Recognition Letters*, 125:85–90.

Alagu, S and K Bhoopathy Bagan. 2019. Acute lymphoblastic leukemia diagnosis in microscopic blood smear images using texture features and svm classifier.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417, Springer.

Chollet, François et al. 2015. Keras. https://github.com/fchollet/keras.

Cui, Hua, Xingwang Liu, Lixin Han, and Zefa Wei. 2019. *Road Crack Classification Based on Improved VGG Convolutional Neural Network*, Fuzzy systems and data mining V : proceedings of FSDM 2019. Amsterdam Ios Press.

Dalal, Navneet and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893, IEEE.

Frid-Adar, Maayan, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. 2018. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331.

Gheisari, Soheila, Daniel Catchpoole, Amanda Charlton, Zsombor Melegh, Elise Gradhand, and Paul Kennedy. 2018. Computer aided classification of neuroblastoma histological images using scale invariant feature transform with feature encoding. *Diagnostics*, 8:56.

Goswami, Shubham, Suril Mehta, Dhruva Sahrawat, Anubha Gupta, and Ritu Gupta. 2020. Heterogeneity loss to handle intersubject and intrasubject variability in cancer. *arXiv:2003.03295 [cs, eess]*.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Howard, Andrew G, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Imran Razzak, Muhammad and Saeeda Naz. 2017. Microscopic blood smear segmentation and classification using deep contour aware cnn and extreme machine learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 49–55.

Jha, Krishna Kumar and Himadri Sekhar Dutta. 2019. Nucleus and cytoplasm–based segmentation and actor-critic neural network for acute lymphocytic leukaemia detection in single cell blood smear images. *Medical Biological Engineering Computing*, 58:171–186.

Jung, Changhun, Mohammed Abuhamad, Jumabek Alikhanov, Aziz Mohaisen, Kyungja Han, and DaeHun Nyang. 2019. W-net: A cnn-based architecture for white blood cells image classification. *arXiv:1910.01091 [cs, eess, q-bio]*.

Karami, Ebrahim, Siva Prasad, and Mohamed Shehata. 2017. Image matching using sift, surf, brief and orb: Performance comparison for distorted images. *arXiv:1710.02726 [cs]*.

Kashif, Muhammad, Thomas M. Deserno, Daniel Haak, and Stephan Jonas. 2016. Feature description with sift, surf, brief, brisk, or freak? a general question answered for bone age assessment. *Computers in Biology and Medicine*, 68:67–75.

Kassani, Sara Hosseinzadeh, Peyman Hosseinzadeh Kassani, Michal J. Wesolowski, Kevin A. Schneider, and Ralph Deters. 2019. A hybrid deep learning architecture for leukemic b-lymphoblast classification.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

LaRussaA. 2015. Childhood blood cancer facts and statistics | leukemia and lymphoma society.

Liu, Peizhong, Jing-Ming Guo, Kosin Chamnongthai, and Heri Prasetyo. 2017. Fusion of color histogram and lbp-based features for texture image retrieval and classification. *Information Sciences*, 390:95–111.

Lowe, G. 2004. Sift-the scale invariant feature transform. *Int. J*, 2:91–110.

Mishra, Sonali, Banshidhar Majhi, and Pankaj Kumar Sa. 2019. Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection. *Biomedical Signal Processing and Control*, 47:303–311.

Mishra, Sonali, Banshidhar Majhi, Pankaj Kumar Sa, and Lokesh Sharma. 2017. Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection. *Biomedical Signal Processing and Control*, 33:272–280.

Naufa, NN and V Sajith. 2020. Automatic leukemia detection in human blood sample based on microscopic images using machine learning.

Ojala, Timo, Matti Pietikäinen, and David Harwood. 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29:51–59.

Patil, Sandeep Baburao and G. R. Sinha. 2016. Distinctive feature extraction for indian sign language (isl) gesture using scale invariant feature transform (sift). *Journal of The Institution of Engineers (India): Series B*, 98:19–26.

Pietikäinen, Matti and Guoying Zhao. 2015. Two decades of local binary patterns. *Advances in Independent Component Analysis and Learning Machines*, pages 175–210.

Putzu, Lorenzo, Giovanni Caocci, and Cecilia Di Ruberto. 2014. Leucocyte classification for leukaemia detection using image processing techniques. *Artificial intelligence in medicine*, 62(3):179–191.

Rawat, Jyoti, Annapurna Singh, H. S. Bhadauria, Jitendra Virmani, and J. S. Devgun. 2017a. Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers. *Multimedia Tools and Applications*, 76:19057–19085.

Rawat, Jyoti, Annapurna Singh, Bhadauria HS, Jitendra Virmani, and Jagtar Singh Devgun. 2017b. Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. *Biocybernetics and Biomedical Engineering*, 37:637–654.

Rehman, Amjad, Naveed Abbas, Tanzila Saba, Syed Ijaz ur Rahman, Zahid Mehmood, and Hoshang Kolivand. 2018. Classification of acute lymphoblastic leukemia using deep learning. *Microscopy Research and Technique*, 81:1310–1317.

Simonyan, Karen and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Srivastava, Divya, Rajitha Bakthula, and Suneeta Agarwal. 2018. Image classification using surf and bag of lbp features constructed by clustering with fixed centers. *Multimedia Tools and Applications*, 78:14129–14153.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

Teuwen, Jonas and Nikita Mariakav. 2020. *Convolutional Neural Network*, Handbook of medical image computing and computer assisted intervention. Academic Press.

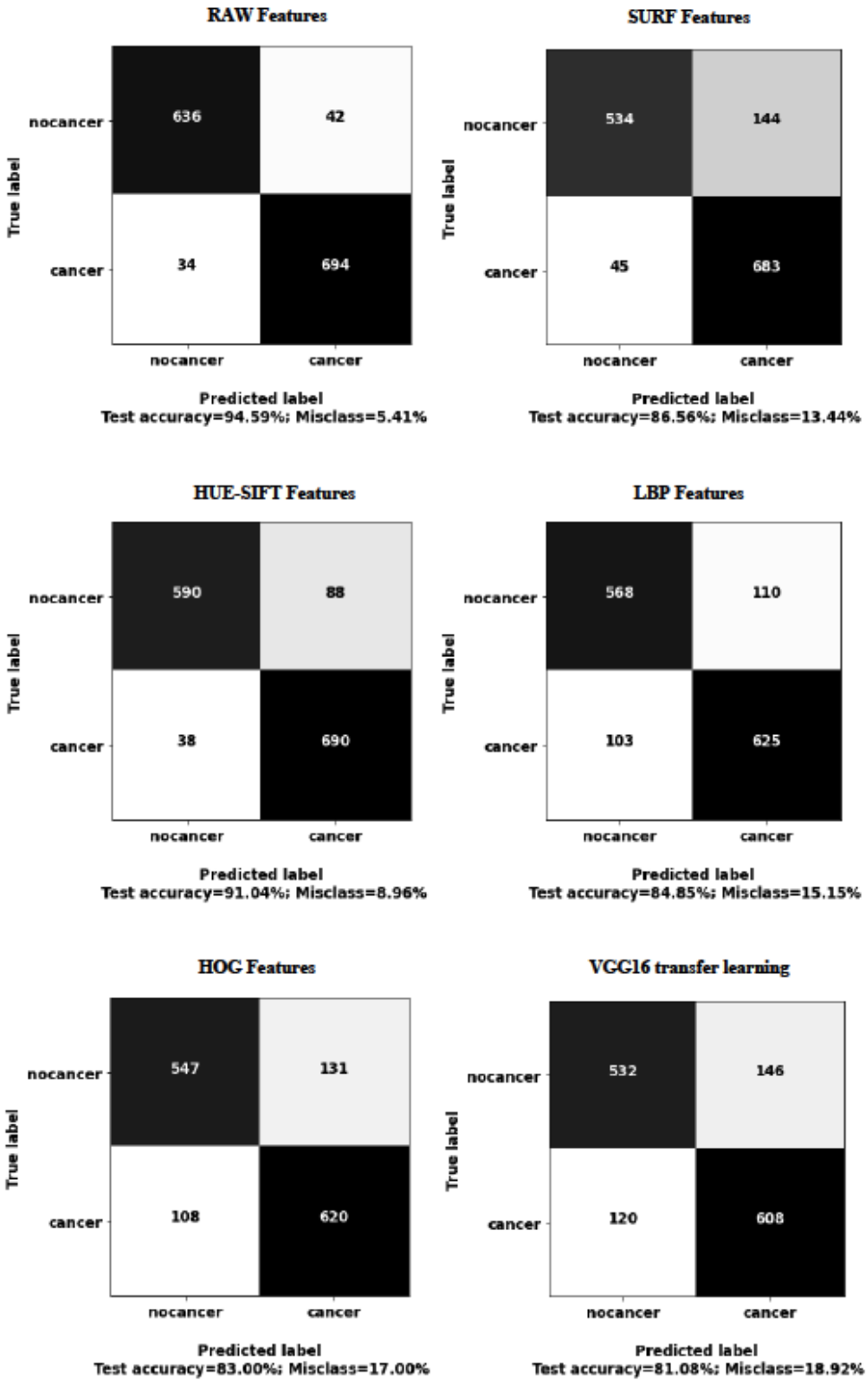
Thanh, T. T. P., Caleb Vununu, Sukhrob Atoev, Suk-Hwan Lee, and Ki-Ryong Kwon. 2018. Leukemia blood cell image classification using convolutional neural network. *International Journal of Computer Theory and Engineering*, 10:54–58.

Tuba, Eva, Ivana Strumberger, Nebojsa Bacanin, Dejan Zivkovic, and Milan Tuba. 2019. Acute lymphoblastic leukemia cell detection in microscopic digital images based on shape and texture features. *In: Tan Y., Shi Y., Niu B. (eds) Advances in Swarm Intelligence. ICSI 2019. Lecture Notes in Computer Science*, 11656:142–151.

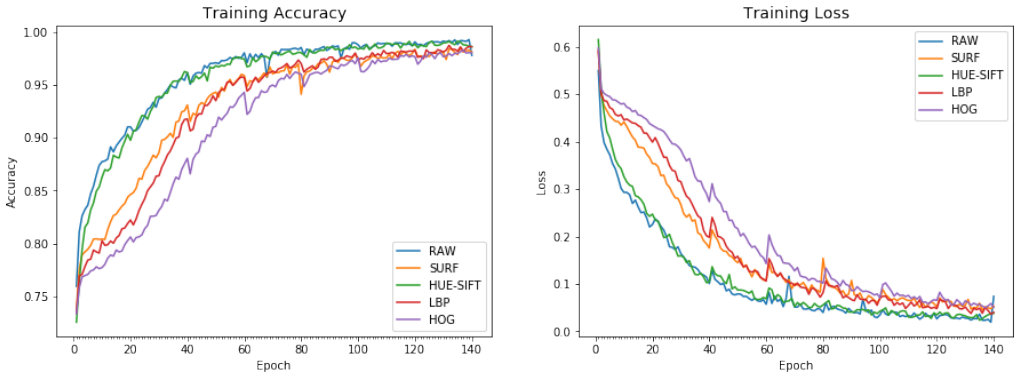
Vijayalakshmi, A et al. 2019. Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, pages 1–21.

Vogado, Luis H.S., Rodrigo M.S. Veras, Flavio. H.D. Araujo, Romuere R.V. Silva, and Kelson R.T. Aires. 2018. Leukemia diagnosis in blood slides using transfer learning in cnns and svm for classification. *Engineering Applications of Artificial Intelligence*, 72:415–422.

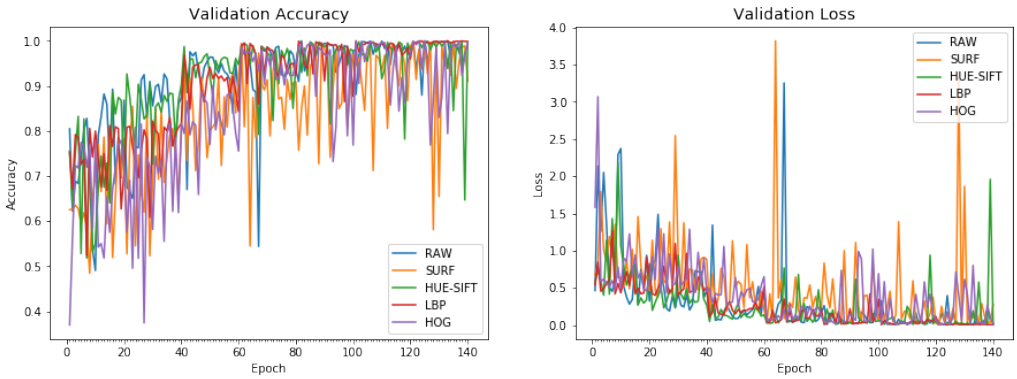
## Appendix A: Confusion matrix for each feature Descriptor



## Appendix B: Training accuracy and loss for each features descriptor trained on a shallow net




## Appendix C: Validation accuracy and loss for each features descriptor trained on a shallow net




## Appendix D: Link to python code used to generate experimental results in this study

https://github.com/mstakale/master_thesis_2020

## Appendix E: Confusion matrix

**Table 1**
Structure of confusion matrix to evaluate performance of classifier

| Classifier Output | Healthy Cells | Cancer Cells |
|---|---|---|
| Healthy cells | TN | FN |
| Cancer cells | FP | TP |