Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2, 682-707
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i2.4570

Optimizing the combination of feature extraction and classifiers in supervised classification of cells images

Xhoena Polisi Duro[1,2*], Arban Uka[2], Griselda Alushllari[2], Albana Ndreu Halili[3], Dimitrios A. Karras[2], Nihal Engin Vrana[4]
[1]Department of Informatics, "Fan S. Noli" University, Korça, Albania; xpolisi@epoka.edu.al (X.P.D.).
[2]Department of Computer Engineering, EPOKA University, Tirana, Albania; auka@epoka.edu.al (A.U.) galushllari@epoka.edu.al (G.A.) dkarras@epoka.edu.al (D.A.K.)
[3]Department of Medicine, Western Balkans University, Tirana, Albania; albana.halili@wbu.edu.al (A.N.H.)
[4]Spartha Medical, Strasbourg, France; evrana@sparthamedical.eu (N.E.V.)

Abstract: Developments in the medical field have opened the opportunity to conduct analyses on a personalized patient level. One of the important analyses that can be conducted is the cellular response to engineered materials, and the most appropriate non-invasive methods are imaging. These images of the cells are unstained brightfield images, as they are acquired from multiparametric microfluidic chambers in the presence of biomaterials and fluids that can change the optical path length over time as the cells' health state is monitored. These experimental conditions lead to an image dataset with unique illumination, texture, and noise spectrum. This study explores the optimization of supervised cell classification by combining feature extraction architectures and machine learning classifiers, with a focus on applications in biomaterial risk assessment. Brightfield microscopy images of three cell types (A549, BALB 3T3, and THP1) were analyzed to evaluate the impact of Inception V3, Squeeze Net, and VGG16 architectures paired with classifiers including KNN, Decision Tree, Random Forest, AdaBoost, Neural Networks, and Naïve Bayes. Dimensionality reduction using Information Gain was applied to improve computational efficiency and accuracy. Butterworth filters with varying parameters were used to balance the enhancement of image features and noise removal, improving classification performance in certain cases. Experimental results demonstrate that the VGG16 architecture, when paired with Neural Networks, achieves higher classification accuracy as measured by different metrics. The improved accuracy when using Butterworth filters compared to the unfiltered dataset and the differences between various Butterworth filters indicate the importance of optimizing filter parameters for these types of images.

Keywords: *Biomaterial risk assessment, Cell image classification, Classifier, Feature extraction, Personalized medicine, Supervised classification.*

1. Introduction

Biomedical research covers a broad spectrum of areas that often are highly interdisciplinary by integrating molecular biology [1] bioelectronics engineering [2] genetics [3] biomaterials [4, 5] clinical medicine [6] and data science [7]. The major objective is to understand biological processes and disease mechanisms, to improve diagnostics and intervention, to develop effective treatments, to design and synthesize materials that interact safely with the biological systems. Over the last two decades, a range of advancements in biomaterial engineering have significantly enhanced the field, making it possible to conduct patient evaluations and treatments at a more personalized level. Major drives include: i) high-resolution 3D bioprinting, ii) novel nanofabrication techniques, iii) development

History: Received: 9 December 2024; Revised: 28 January 2025; Accepted: 28 January 2025; Published: 3 February 2025
* Correspondence: xpolisi@epoka.edu.al

of functionalized bioactive coatings, iv) stimuli responsive hydrogels, v) advanced imaging [8]. The control of the porosity and the pore size on the nanofibers are important parameters in biomaterial-cell interaction and these can both be fabricated with high control aiming to mimic biological systems [9, 10]. These three-dimensional structures can be constructed using nanofibers produced by electrospinning and the whole serves as an extracellular matrix that provides mechanical support for the biological samples that can be studied [11]. These structures facilitate the evaluation of the cell interaction with nano-structured and micro-structured surfaces to further study the cellular response to engineered materials [12]. The development of bioactive coatings that are patient-specific [13, 14] the novel use of stimuli-responsive hydrogels and the development of high-resolution microCT imaging help understand better the cell-material interaction at the cellular level.

Often the material that needs to be studied is a biomaterial whose toxicity information at the personalized level is essential [15]. Having the most biocompatible biomaterial evaluated in this form is a new possibility for all patients in need of implants. To determine this, microscopy has to be used for a large number of images for experiments conducted mostly on multiparametric microfluidic chips over the course of a few weeks [16]. The cellular response to engineered surfaces and biomaterials at different concentrations is often exhibited in electrochemistry-based measurements such as impedance and chronoamperometric measurements [17]. The need to conduct measurements that are affected by free charges or ions means that one cannot use staining for contrast enhancement thus limiting the contrast. At the same time the cells reside in hydrogels that mimic tissue behaviour and whose index of refraction varies depending on their relative hydration level. The spatial variation of the index of refraction often results in images that mimic speckle patterns. Further details on the challenges of microscopy in these settings are provided in Uka, et al. [18].

These limitations in microscopy can be overcome by employing: i) hardware-based methods such as using incoherent light, rotating diffusers [19]; ii) software-based methods including wavelet transform denoising [20] gaussian filters; iii) hybrid methods by combining hardware and software methods and iv) image reconstruction algorithms [21]. In the case of selecting a biomaterial for an implant, medical practitioners need to test a few samples (blood, saliva, different cell types etc) from a patient in contact with different biomaterials at different concentrations. The microscope images need to undergo both qualitative (classification) and quantitative (area of cells, perimeter of cells, cell counting etc) evaluation [8, 22-24] and these results help deduce the toxicity of a biomaterial [18]. All the insights that we learn are focused particularly in assessing biomaterial toxicity level by studying changes in the morphological features of different immune cells.

Analyzing the changes in the morphological features of same cells constitutes an intraclass cell classification task, i.e. evaluating fine grained to extreme differences on the components of a cell such as the cytoplasm. All the images that are acquired using brightfield microscopy in the microfluidic chamber settings - including different cell types - have common features including illumination, texture, dimensionality, and noise spectrum. Interclass cell image classification would require the identification of slightly high differences in the images [25, 26]. Overall, the intraclass and interclass brightfield cell image classification constitutes an excellent and robust platform for benchmarking the performance of deep learning models. The techniques may include both supervised [15, 27] and unsupervised methods [28, 29]. This study focuses on classification of unstained images of three distinct cell lines: A549 (human lung carcinoma), Balb 3T3 (fibroblast) and THP1 (monocyte) prior to their exposure to the biomaterials. The cell images are acquired by Dr. Vrana's group (a co-author of this study) using a brightfield microscope, which has a lower resolution and contrast compared to advanced imaging systems.

Inception V3, SqueezeNet and VGG 16, all pre-trained using the ImageNet dataset for object detectetion, are the selected architectures to extract the image features. These three architectures have been commonly reported in the literature and their selection was done based on their fast and accurate performance in this specific dataset. Inception V3 architecture extracts higher dimensional features by

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

balancing depth and computational cost of different size filter used at the same layer Szegedy, et al. [30] and Tamilarasi and Gopinathan [31]. Wang, et al. [32] reported a sensitivity of 95.41% when classifying pulmonary chest X-ray images. SqueezeNet architecture can achieve a high image classification accuracy even though it is strongly compressed and uses few parameters compared to other architectures [33]. This architecture and its improved versions have achieved CT scan classification accuracy from 93.2%-95.8% [34]. VGG16 architecture is a simple 16 layered architectures that uses 3x3 convolution kernels and can be further changed to increase its accuracy [35].

Then, the extracted features are classified using various machine learning algorithms such as: K Nearest Neighbors (KNN), Decision Tree, Random Forest, AdaBoost, Simple Neural Network and Naïve Bayes. Research shows that the neural networks classifiers have a better performance compared to other models [36].

In order to reduce the complexity and lower the number of features, several methods can be used such as Principal Components Analysis (PCA), Information Gain, combination of both- a hybrid form [37-39]. The efficiency and the performance of the model is evaluated when: a) all extracted features are and b) 10 most important features are selected through Information gain. [40]. During the evaluation process different metrics such precision, recall, F1 score, accuracy etc. are used.

In evaluating for a robust framework for cell image classification several filters can also be used [41] as they can significantly improve the quality of the features by providing a balance between noise reduction and feature preservation. The use of the Butterworth filters with the objective of improving classification may strongly depend on the dataset, its quality, texture and noise spectrum and careful analysis would be needed in our cell image dataset. Minoshima, et al. [42] have optimized the cut off frequencies of the Butterworth filters for brain images, and in other works where optimal combination of filters including Butterworth is reported [43]. This work aims to find a robust method that is fast and accurate that can be easily implemented in the future in a low-cost, portable biomaterial personalized testing device. After analyzing the optimal architecture-classifier combination we evaluate the effect of the Butterworth filters with different parameters in the image classification.

2. Materials

2.1. Image Features

The classification is done on three different cell image datasets. A549, Balb 3T3, and THP-1 cells, each representing a distinct tissue or immune function, exhibit different baseline morphologies that can be used to accurately distinguish them from each other prior to biomaterial exposure.

A549 cells, derived from human lung epithelial carcinoma, display a polygonal, cobblestone-like morphology, typical of epithelial cells. These cells tend to grow in a sheet-like manner, with tight cell-to-cell junctions, making them appear compact and well-ordered under high-resolution microscopy [44]. Balb 3T3 cells, a fibroblast cell line, show an elongated, spindle-like shape, characteristic of fibroblasts. These cells exhibit a scattered growth pattern with elongated processes extending from the cell body. Their morphology is key in distinguishing them as fibroblast-like cells that contribute to connective tissue formation [45]. THP1 cells, a monocyte cell line, exhibit a round, non-adherent morphology when in their undifferentiated state. THP-1 cells float in suspension, unlike A549 and Balb 3T3, and do not adhere to culture surfaces. Upon differentiation into macrophages, they become adherent and take on an irregular, spread-out shape, which is often used as a marker of macrophage activation [46].

These three different cells all have different cell features, and it is easier to distinguish THP1 from the others when they are not living in clusters due to their round shape (see Figure 1). However, BALB 3T3 and A549 can be easily confused, especially when the number of cells in an image is very high. The images are taken using a brightfield microscope, having low contrast, different brightness, and non-uniform background illumination. All these factors may affect the identification of cell features, especially at images where the edges of the cells are very difficult to distinguish from the background.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

This would significantly decrease the accuracy of the cell identification and classification.

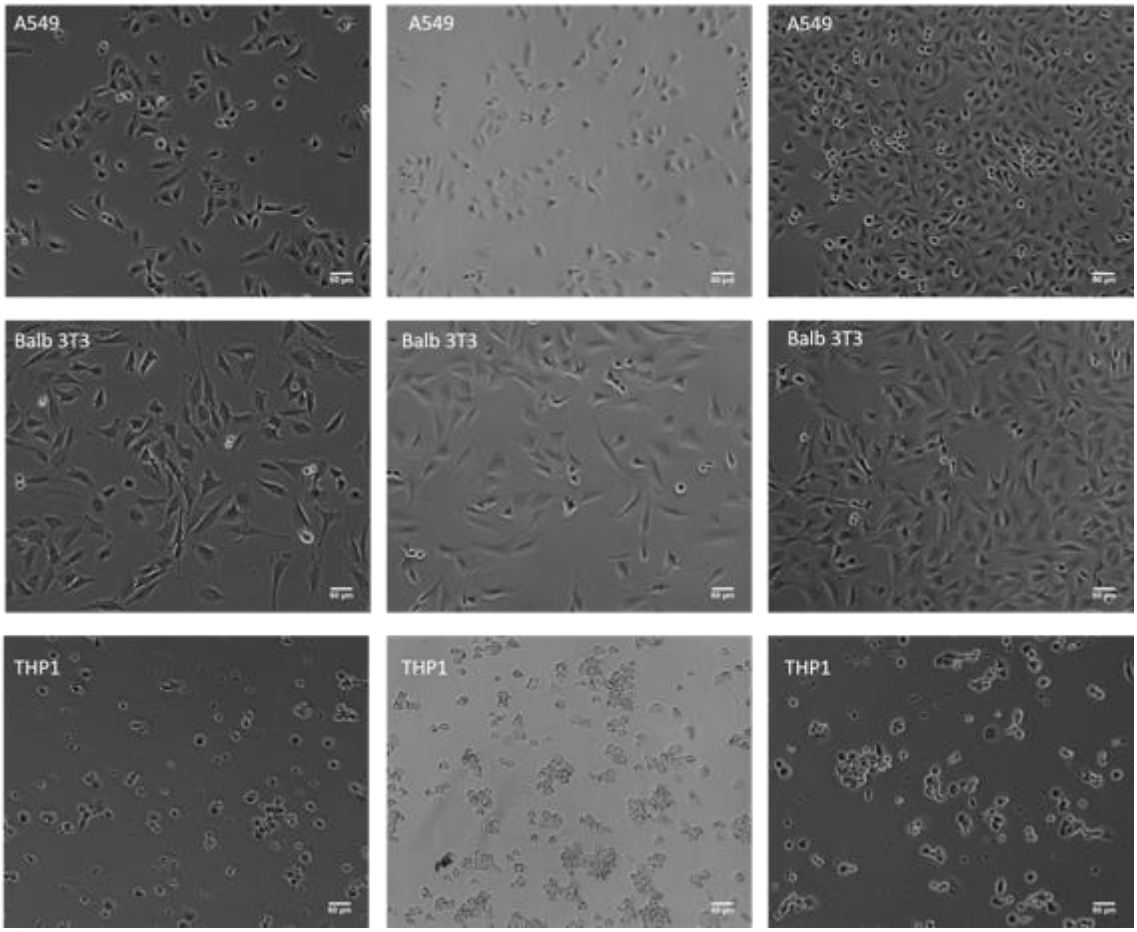

Figure 1.
Different cell image features for A549, BALB 3T3 and THP1.

2.2. Dataset 1 Details

The dataset consists of 252 images in total, of size 1024 x 1280 pixels, taken at day one at three different types of cells: A549, BALB 3T3 and THP1. There are 84 images per each type of cell, divided into 56 images for training and 28 images for testing, 66% and 34% respectively. So, the training dataset consists of 168 images in total, and the testing dataset has 84 images.

2.3. Dataset 2 Details

The number of original images is low and the image size is very large compared to cell size. The original images are cropped to smaller ones down to 256x256 pixels. The newly formed dataset is cleaned by removing empty crops (only background). After cleaning, there are 1580 images per each type of cell, divided into 1052 (66%) and 528 (34%) images into training and testing dataset respectively. So, in total the training dataset consists of 3156 images and the testing dataset consists of 1584 images. Some Balb 3T3 cells have a large cytoplasm (when the confluency is low) that can be represented by image sections larger than 128x128 pixels and this the reason why the smallest cropped image used is 256x256.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

2.4. Butterworth Filter

Butterworth filters are both low-pass and high pass filters that are widely used for image enhancement [47, 48]. These filters are applied in frequency domain and are known to effectively reduce noise while preserving important information. Their usage in these types of cell images, are very useful since they can enhance cytoplasm features, sharpen the image without the blurring effect, making it easier for the cells to be detected, especially on the images that have a low contrast compared to the background (see Figure 2,3). The passband and the stopband frequencies can be changed to find the optimal filter values.

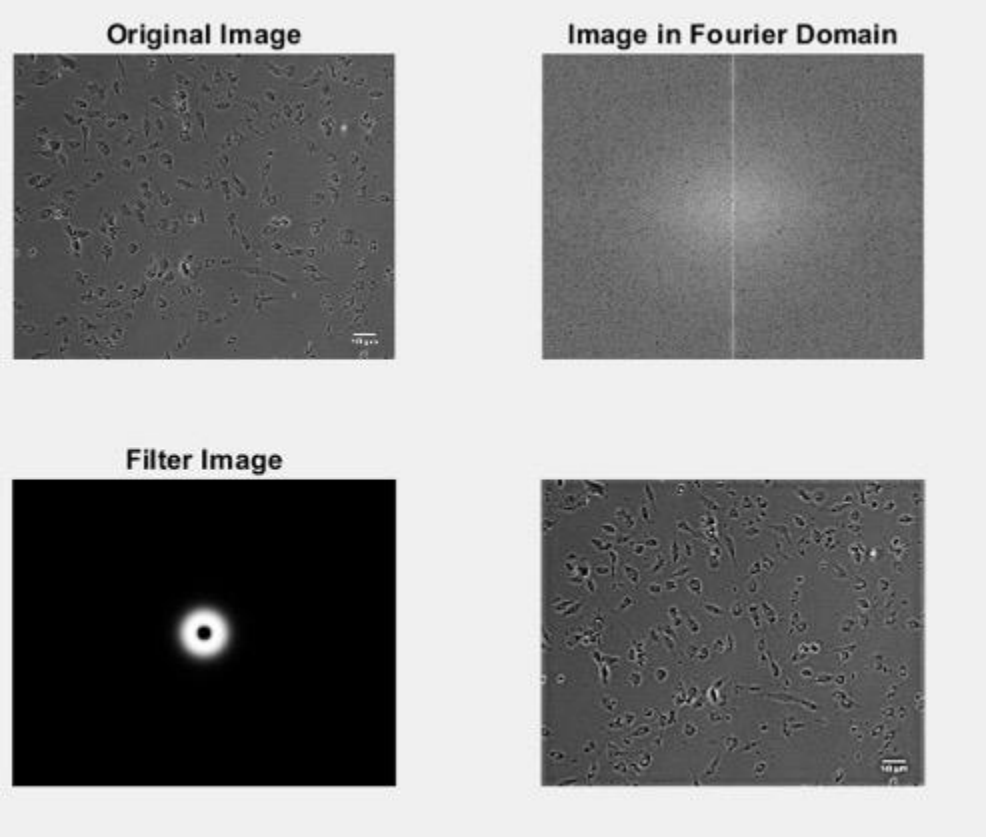

Figure 2.
Filter spectrum for the Bandpass filter (up) and an example of the usage of Butterworth filter in D1.

The order of the filter is another parameter that affects how much we want to smooth out fine-grain details or enhance the edges and cell contours. In this work, three types of filters BW1(lowpass=30, highpass=120, order=4), BW2 (lowpass=30, highpass=150, order=4), BW3 (lowpass=50, highpass=150, order=4) will be used to check how the accuracy changes when a preprocessing step is applied to the original images (Dataset 1).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

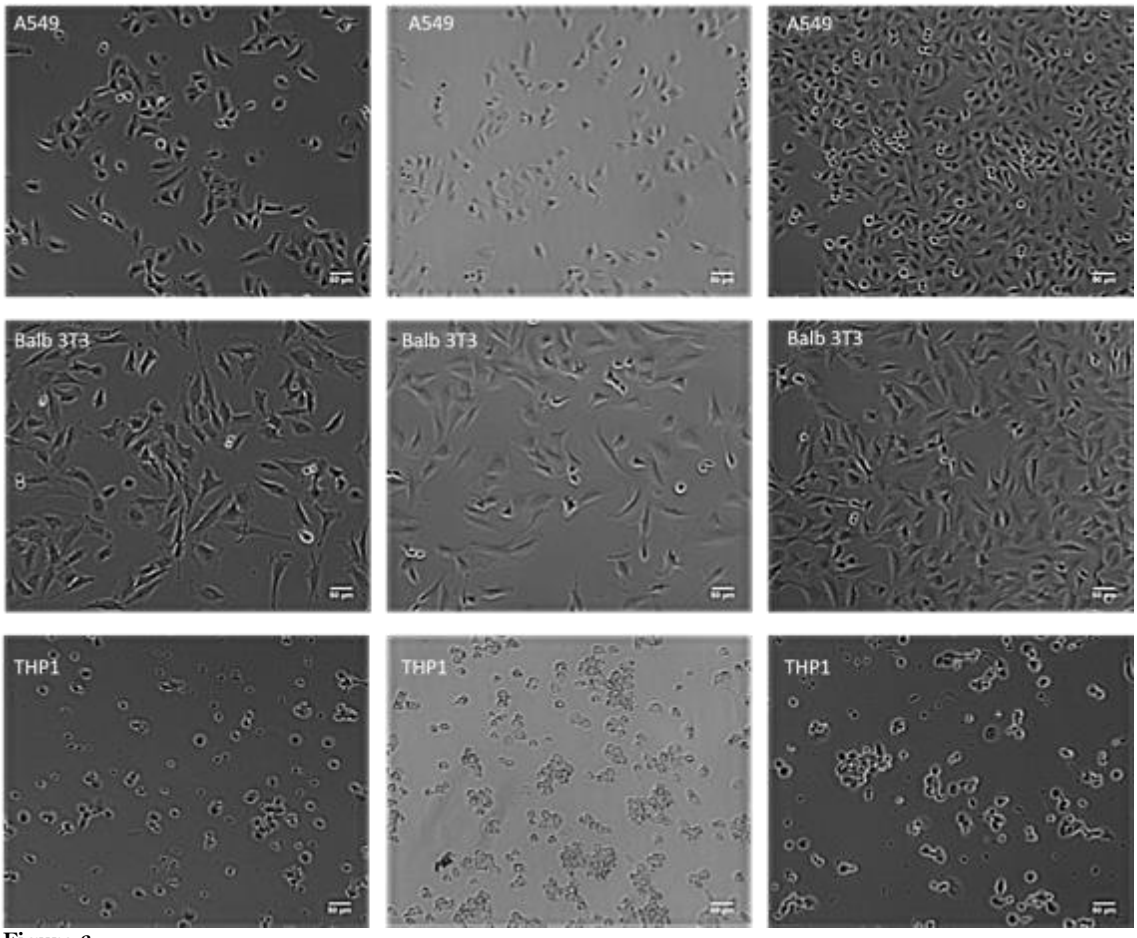

Figure 3.
Cell images for A549, BALB 3T3 and THP1 after applying Butterworth filters.

3. Methods

3.1. Feature Extraction

The key cell differences between these cell types can be summarized as: a) A549 cells have a polygonal shape, tight cell junctions, cobblestone appearance; b) Balb 3T3 cells have a spindle-like, elongated shape, scattered growth with extended processes; c) THP1 cells are round, non-adherent (in monocytic form). A higher classification performance means that the algorithm can detect these on the extracted image features. Dhal and Azad [49] provide a deep analysis regarding feature selection and feature extraction methods, based on related research. The extraction of image features is done by using three different architectures: Inception V3, SqueezeNet and VGG 16. They are all pre trained on ImageNet, a very large dataset consisting of more than 14 million images, 20.000 categories that is used in object recognition research.

Inception V3 [30] utilizes inception modules to extract image features at different scales. These modules apply convolutions of different sizes (1x1, 3x3, 5x5) simultaneously, rather than sequentially, in order to extract the features. This aspect is very effective when analyzing images of varying features scales, and this is why Inception V3 architecture performs well on cell images analysis. Inception V3 has been successfully used in medical image analysis, including cancer cell classification [50, 51] and pathology image recognition [52-55] due to its efficiency in feature extraction and classification accuracy. The total number of extracted feature vector sizes per image is 2048.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

SqueezeNet [33] is a very efficient architecture with a compact design, that reaches high accuracy with way less parameters compared to other architectures. It uses fire modules to balance accuracy with computational efficiency. The fire modules provide two different processes "squeezing" and "expanding. The first one compresses the input channels with 1x1 convolutions, while the second one increases the input channels by combining 1x1 and 3x3 filters. Its compressed nature, mixed with the ability to extract important features and produce high results from them, makes it ideal for usage in a portable, low-cost device. This architecture has been used for tumor cell image classification [56-58] and other object classification [59]. The total number of extracted features per image is 1000.

VGG-16 [60] architecture is deeper than the other mentioned ones, and requires more time during the training and testing process. This is due to the fact that it has 16 layers, primarily composed of 3x3 convolutional filters, but the number can be adjusted to reach higher accuracy. Because of its nature, it is able to extract hierarchical features from images and recognize complex patterns, which is why VGG-16 has been extensively used in medical imaging tasks [61] including cell classification [62] tumor detection [63] and other diagnostic tasks [64]. While VGG-16 is very powerful in feature extraction, the number of features extracted per image is high, 4096 to be precise, which may not be ideal in terms of computational power.

As the number of features extracted from these three models is high, dimensionality reduction using Information Gain is applied to select top 10 relevant features per each model. Information Gain finds these features by evaluating each feature contribution to the classification, and then selects those that can distinguish each class better [40, 65]. Applying dimensionality reduction to 10 most relevant features is important for several reasons such as: i) computational efficiency, since a smaller feature dataset reduces the computational power, making it feasible for implementation in a portable, low- cost device; ii) reduced risk of overfitting and iii) faster processing and real time evaluation, which is essential in these types of devices.

3.2. Classifiers

The classifiers selected in this work are KNN, Decision Tree, Random Forests, AdaBoost, Neural Network and Naïve Bayes.

K-Nearest Neighbours [66] is a simple yet effective classification algorithm that classifies instances based on the closest training examples in the feature space. For each new cell image, the algorithm identifies the "k" nearest neighbours in the training data and assigns the majority class among those neighbours as the predicted class. The choice of "k" can affect the algorithm's performance, small values of "k" can lead to sensitivity to noise, while larger values can result in smoother boundaries between classes. It can be very effective for real time applications, but often struggles in high dimensionality data.

Decision Tree [67] is a conditional classifier, where features are shown as nodes, and the conditional rules, shown as edges, connect the nodes until the predicted class label, and they also decide the flow of the algorithm until the final decision. They offer a visual representation of the combination of features per each final class decision. Due to their conditional rules, they are a good choice for problems with non-linear features, but may suffer from overfitting.

Random Forest [68] uses multiple decision trees on its core, in order to improve the performance. It tries to overcome the problem of overfitting in Decision Trees, by calculating the average of predictions of the trees, making it more robust to noisy data or small datasets. During the training process, it selects a random subset from the training data and uses it to train each tree in the forest. Every tree generates a random subset of features, and the results are aggregated in the end. By doing so, the model is able to train on each features of the dataset, proving to be very effective when classifying noisy and high dimensionality data.

AdaBoost [69] Adaptive Boosting is an ensemble learning method that tries to increase the accuracy of the weak classifiers by putting more emphasis on the hard to classify objects or instances.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

The training process is done in a group of classifiers sequentially and for each iteration, it gives more weight to the wrongly classified objects or instances from the previous step. The final result is a combination of all the classifiers, giving more weight to the ones that had higher accuracy on the hard to classify instances.

Neural Network [70] uses multiple layers of neurons in order to extract complex. The neurons are connected together in every successive layer, where each neuron-connection represents a feature transformation. The training process is done in a hierarchical model, which often gives a high accuracy, to the cost of computational power. They are used for various tasks in image analysis.

Naïve Bayes [71] classifiers is based on the Bayes Theorem and on the estimation that the features extracted are conditionally independent from the class label. Certainly, this assumption may not be true, however, its performance is quite effective in different applications. The algorithm calculates the posterior probability for each class. The class that has the highest value is chosen to be the final predicted class label. This simple, yet effective classifier has proven to be a good selection for datasets containing high number of features.

3.3. Evaluation Metrics

The performance of each classifier during the training process is evaluated through a 5-fold and 10-fold cross validation method. This method divides the training dataset into 5 or 10 subsets. Then, it randomly selects a subset to be the testing set, while the rest are used for training. Several studies have studied the effect of the cross validation [72] and there are reports that there are optimal values for k in the k-fold cross validation that directly affect the accuracy [73]. Nevertheless, cross validation method ensures that each subset of the data is used both as a training/testing set, making it very robust and providing more reliable information. During the testing process, a blind dataset is used to make sure to check how the classifiers perform in unseen data. The performance evaluation metrics that are used are accuracy, precision, recall (sensitivity), F1 score, Area under the Curve (AUC), and MCC.

Accuracy (1) is calculated as the ratio of the correctly classified instances or object over the whole dataset. Although it is a fundamental metric, it may not be the best indicator for imbalanced datasets, where a dominant class can disproportionately influence the results.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Samples} \quad (1)$$

Precision (2) evaluates the classifier's ability to avoid false positives. It focuses on the proportion of the true positive predictions relative to all positive predictions made by the model.

$$Precision = \frac{True\ Positives}{True\ positives + False\ Positives} \quad (2)$$

Recall (Sensitivity) (3) assesses the classifier's ability to identify true positives. It is the proportion of true positive predictions relative to all actual positives in the dataset.

$$Recall = \frac{True\ Positives}{True\ positives + False\ Negatives} \quad (3)$$

F1 score (4) balances precision and recall, since it is the harmonic mean of precision and recall. It is particularly useful when there is an imbalance between precision and recall, as it provides a single measure of performance that considers both false positives and false negatives.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

The Matthews Correlation Coefficient (MCC) (5) is a robust metric that can handle imbalanced datasets, with values ranging from -1 to +1, where +1 indicates perfect predictions and -1 shows total disagreement. The value 0 signifies a random prediction.
Due to its robustness and interpretability, it can be applied in different classification problems.

$$MCC = \frac{True\ Positives * True\ Negatives - False\ Positives * False\ Negatives}{\sqrt{A * B * C * D}} \quad (5)$$

$$A = True\ Positives + False\ Positives$$
$$B = True\ Positives + False\ Negatives$$
$$C = True\ Negatives + False\ Positives$$
$$D = True\ Negatives + False\ Negatives$$

Area Under the Curve (AUC) measures the classifier's ability to distinguish between classes across different thresholds. AUC ranges from 0.5 (random guessing) to 1.0 (perfect classification). A summary of all parameters involved in this study is shown in Table 1, together with their corresponding abbreviations

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

Table 1.
Summary of all the parameters of this study with their abbreviations.

Datasets		
Name	**Abbreviation**	
Dataset 1	D1	• Total 252 images • 84 per cell • Size 1024*1280 • Training 168 images • Testing 84 images
Dataset 2	D2	• Total 4740 images • 1580 per cell • Size 256 x 256 • Training: 3156 images • Testing: 1584 images
Butterworth filters		
Butterworth filter 1	BW1	• Lowpass: 30 • Highpass:120 • Order: 4
Butterworth filter 2	BW2	• Lowpass: 30 • Highpass:150 • Order: 4
Butterworth filter 3	BW3	• Lowpass: 50 • Highpass:150 • Order: 4
Feature extraction		
Name	Abbreviation	
Inception V3	IV3	Nr of features: 2048
SqueezeNet	SN	Nr of features: 1000
VGG-16	VGG	Nr of features: 4096

Classifiers					
K-Nearest Neighbor KNN	Decision Tree DT	Random Forest RF	AdaBoost AB	Neural Network NN	Naïve Bayes NB
Evaluation Metrics					
Precision PREC	Classification Accuracy CA	Recall REC	Area Under Curve AUC	F1 Score F1	Mathew correlation coefficient MCC

4. Results

4.1. Dataset 1 Results

Below are shown the overall performance comparison of three feature extraction architectures—Inception V3 (IV3), SqueezeNet (SN), and VGG-16 —combined with six classifiers (KNN, Decision Tree, Random Forest, AdaBoost, Neural Network, and Naïve Bayes) for the classification of three different cell types (A549, Balb 3T3, THP1). The results reported in Table 2 are obtained when all features are used in the training and testing. Each classifier-architecture pair is evaluated using multiple performance metrics, including AUC, CA, F1 score, Precision (PREC), Recall (REC), and MCC, under three validation schemes (5-fold, 10-fold cross-validation) and an independent test set. This detailed evaluation enables a robust assessment of both the discriminative power of extracted features and the generalization ability of the employed classifiers. High AUC values—often converging to or exceeding

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

0.99—suggest excellent discrimination capabilities, while consistently high CA, F1, PREC, REC, and MCC scores across training and test conditions indicate stable and reliable predictive performance.

Table 2.
Comparative performance of six classifiers using three feature extraction architectures for three cell type classification for D1.

Training Process

Classifier /Model		AUC		CA		F1		PREC		REC		MCC	
		5F	10F	5F	10F	5F	10F	5F	10F	5F	10F	5F	10F
KNN	IV3	0.994	0.999	0.97	0.976	0.97	0.976	0.97	0.977	0.97	0.976	0.955	0.964
	SN	0.999	0.999	0.964	0.976	0.964	0.976	0.965	0.977	0.964	0.976	0.947	0.965
	VGG	0.999	0.999	0.976	0.976	0.976	0.976	0.977	0.977	0.976	0.976	0.964	0.964
DT	IV3	0.928	0.908	0.881	0.863	0.881	0.863	0.883	0.864	0.881	0.863	0.822	0.795
	SN	0.94	0.93	0.905	0.887	0.905	0.887	0.907	0.888	0.905	0.887	0.858	0.83
	VGG	0.928	0.934	0.887	0.893	0.887	0.893	0.888	0.894	0.887	0.893	0.831	0.84
RF	IV3	0.992	0.997	0.946	0.964	0.947	0.964	0.952	0.965	0.946	0.964	0.922	0.947
	SN	0.998	0.997	0.964	0.964	0.964	0.964	0.965	0.964	0.964	0.965	0.947	0.947
	VGG	0.994	0.996	0.946	0.958	0.946	0.958	0.946	0.958	0.946	0.958	0.92	0.938
AB	IV3	0.924	0.902	0.899	0.869	0.899	0.869	0.9	0.871	0.899	0.869	0.848	0.804
	SN	0.933	0.906	0.911	0.875	0.911	0.876	0.911	0.878	0.911	0.875	0.866	0.813
	VGG	0.906	0.915	0.875	0.887	0.876	0.887	0.878	0.889	0.875	0.887	0.813	0.831
NN	IV3	0.997	0.996	0.958	0.964	0.958	0.964	0.958	0.965	0.958	0.964	0.938	0.947
	SN	1	1	0.982	0.994	0.982	0.994	0.982	0.994	0.982	0.994	0.973	0.991
	VGG	0.996	0.996	0.964	0.964	0.964	0.964	0.965	0.965	0.964	0.964	0.947	0.947
NB	IV3	0.992	0.994	0.964	0.964	0.964	0.964	0.964	0.964	0.964	0.964	0.946	0.946
	SN	0.989	0.992	0.952	0.964	0.953	0.965	0.956	0.968	0.952	0.964	0.93	0.948
	VGG	-	-	0.905	0.917	0.906	0.917	0.912	0.921	0.905	0.917	0.86	0.877

Testing process

Classifier /Model		AUC	CA	F1	PREC	REC	MCC
KNN	IV3	0.99	0.94	0.94	0.949	0.94	0.916
	SN	0.989	0.964	0.965	0.968	0.964	0.948
	VGG	0.991	0.976	0.976	0.978	0.976	0.965
DT	IV3	0.866	0.821	0.821	0.835	0.821	0.739
	SN	0.918	0.881	0.878	0.905	0.881	0.834
	VGG	0.962	0.929	0.93	0.935	0.929	0.895
RF	IV3	0.997	0.893	0.89	0.919	0.893	0.854
	SN	0.98	0.845	0.848	0.894	0.845	0.791
	VGG	0.999	0.976	0.976	0.978	0.976	0.965
AB	IV3	0.875	0.833	0.833	0.843	0.833	0.755
	SN	0.955	0.94	0.94	0.943	0.94	0.912
	VGG	0.973	0.964	0.964	0.966	0.964	0.947
NN	IV3	0.999	0.988	0.988	0.989	0.988	0.982
	SN	1	1	1	1	1	1
	VGG	1	0.988	0	0.989988	0.988	0.982
NB	IV3	0.989	0.964	0.964	0.968	0.964	0.948
	SN	0.983	0.929	0.929	0.936	0.929	0.896
	VGG	-	0.976	0.976	0.978	0.976	0.965

Comparing the fivefold and ten fold cross validation splitting shows the stability of the classifier when trained and tested. Generally, similar values in 5F and 10F suggest that the model is robust to how the training data is partitioned. Analysis of the results reveals that certain classifier and feature extraction combinations outperform others. Neural Networks demonstrate near-perfect classification performance when paired with SqueezeNet or VGG features, consistently achieving high AUC, CA, and F1 scores and MCC values reflecting nearly flawless predictions (See Figure 4). Similarly, Random Forest and KNN models, when trained on either IV3 (see Figure, 5,6) or VGG-derived features, exhibit

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

notably high overall performance metrics, indicating that well-engineered deep representations can effectively separate cell type classes.

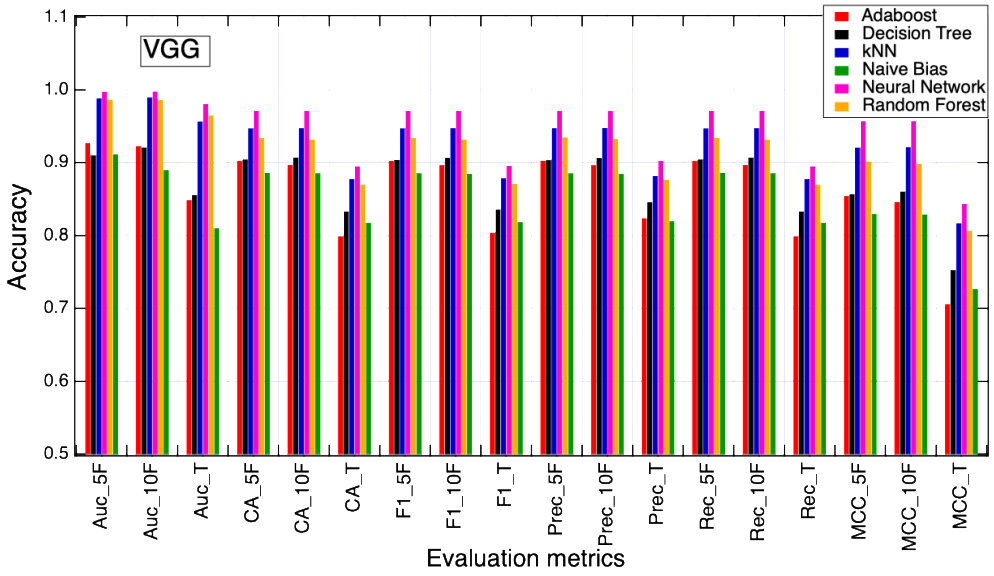

Figure 4.
Accuracy metrics of 5-fold, 10-fold splitting and testing for all classifiers when VGG 16 architecture is used.

In contrast, Decision Trees and AdaBoost, while still producing good results, generally lag behind the top-performing models, suggesting a lesser capacity to fully leverage the richness of deep features or to establish complex decision boundaries required for optimal differentiation among cell types.

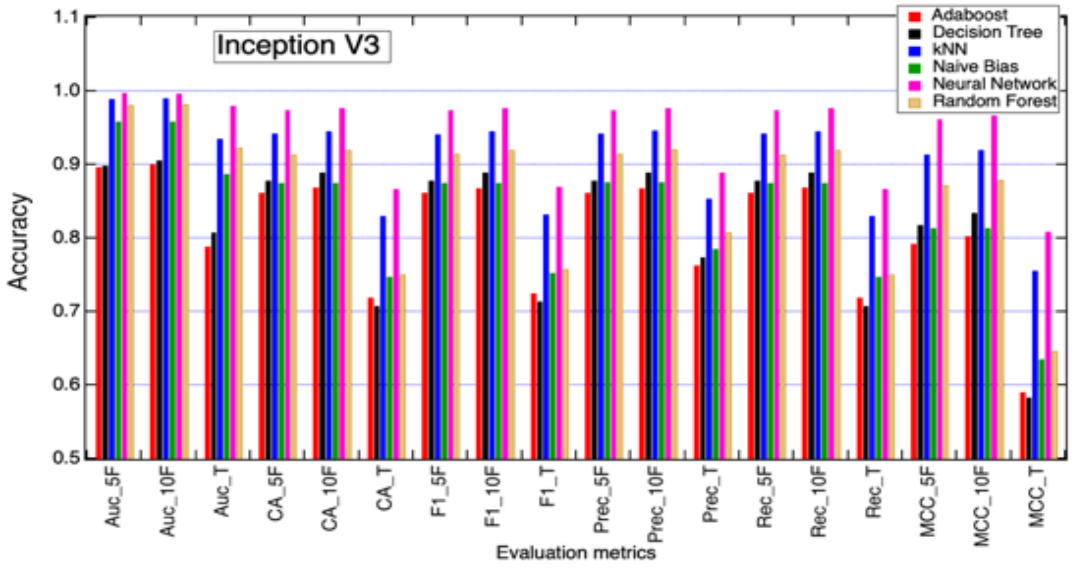

Figure 5.
Accuracy metrics of 5-fold, 10-fold splitting and testing for all classifiers when Inception V3 architecture is used

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

These findings suggest that the most optimal configurations are those that pair robust feature extraction architectures with more sophisticated classification algorithms. In particular, SqueezeNet- and VGG-based feature vectors combined with Neural Networks emerged as the top-tier solutions, delivering superior AUC, CA, and F1 metrics, as well as stable MCC values across cross-validation and independent testing. Such combinations not only highlight the importance of selecting advanced, high-quality feature representations but also underscore the value of choosing classifiers capable of capitalizing on these complex feature sets. In practice, these insights would guide researchers toward leveraging deep-learning-derived features in conjunction with neural network models to achieve maximal accuracy, reliability, and scalability in high-throughput cellular classification tasks.

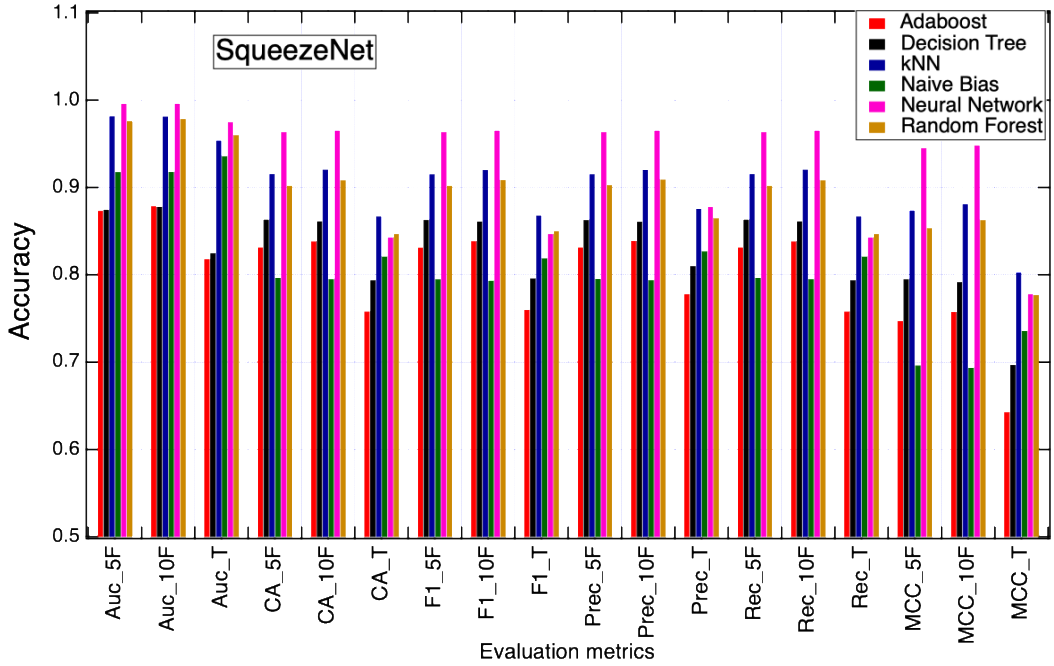

Figure 6.
Accuracy metrics of 5-fold, 10-fold splitting and testing for all classifiers when SqueezeNet architecture is used

Similar to the previous one, the same experiment is repeated by changing the number of features in the training and testing process. Only 10 most relevant features are selected using Information Gain, and the evaluation metrics are shown in Table 3.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

Table 3.
Comparative performance of six classifiers using 10 relevant features obtained using dimensionality reduction from Information Gain for D1.

Training Process													
Classifier /Model		**AUC**		**CA**		**F1**		**PREC**		**REC**		**MCC**	
		5F	10F	5F	10F	5F	10F	5F	10F	5F	10F	5F	10F
KNN	IV3	0.994	0.994	0.97	0.958	0.97	0.959	0.971	0.96	0.97	0.958	0.956	0.938
	SN	0.99	0.99	0.97	0.964	0.97	0.964	0.971	0.964	0.97	0.964	0.955	0.946
	VGG	0.991	0.999	0.946	0.964	0.946	0.964	0.947	0.964	0.946	0.964	0.92	0.946
DT	IV3	0.933	0.934	0.911	0.923	0.91	0.923	0.91	0.923	0.911	0.923	0.866	0.884
	SN	0.948	0.955	0.911	0.935	0.911	0.935	0.911	0.935	0.911	0.935	0.866	0.902
	VGG	0.926	0.925	0.893	0.869	0.892	0.868	0.892	0.868	0.893	0.869	0.839	0.804
RF	IV3	0.933	0.993	0.911	0.958	0.91	0.959	0.91	0.961	0.911	0.958	0.866	0.938
	SN	0.989	0.993	0.929	0.964	0.929	0.964	0.93	0.965	0.929	0.964	0.893	0.947
	VGG	0.992	0.99	0.946	0.935	0.947	0.934	0.947	0.934	0.946	0.935	0.92	0.902
AB	IV3	0.951	0.946	0.935	0.929	0.935	0.929	0.935	0.929	0.935	0.929	0.902	0.893
	SN	0.929	0.951	0.905	0.935	0.904	0.935	0.905	0.936	0.905	0.935	0.857	0.902
	VGG	0.924	0.92	0.899	0.893	0.898	0.893	0.898	0.893	0.899	0.893	0.849	0.84
NN	IV3	0.998	0.998	0.958	0.958	0.958	0.958	0.959	0.959	0.958	0.958	0.938	0.938
	SN	0.998	0.998	0.97	0.976	0.97	0.976	0.971	0.976	0.97	0.976	0.956	0.964
	VGG	0.997	0.998	0.958	0.958	0.958	0.958	0.959	0.96	0.958	0.958	0.938	0.938
NB	IV3	0.996	0.997	0.94	0.946	0.941	0.947	0.941	0.947	0.94	0.946	0.911	0.92
	SN	0.997	0.997	0.97	0.97	0.97	0.97	0.97	0.971	0.97	0.97	0.955	0.956
	VGG	0.996	0.996	0.94	0.952	0.941	0.953	0.944	0.956	0.94	0.952	0.912	0.93

Testing Process							
Classifier /Model		AUC	CA	F1	PREC	REC	MCC
KNN	IV3	0.981	0.952	0.952	0.958	0.952	0.932
	SN	1	0.976	0.976	0.978	0.976	0.965
	VGG	1	0.976	0.976	0.978	0.976	0.965
DT	IV3	0.962	0.94	0.939	0.947	0.94	0.915
	SN	0.898	0.869	0.871	0.875	0.869	0.805
	VGG	0.959	0.929	0.928	0.933	0.929	0.895
RF	IV3	0.996	0.905	0.902	0.922	0.905	0.868
	SN	0.999	0.976	0.976	0.978	0.976	0.965
	VGG	0.999	0.952	0.953	0.958	0.952	0.931
AB	IV3	0.955	0.94	0.94	0.949	0.94	0.916
	SN	0.902	0.869	0.87	0.877	0.869	0.806
	VGG	0.955	0.94	0.941	0.949	0.94	0.914
NN	IV3	0.999	0.964	0.964	0.966	0.964	0.947
	SN	0.998	0.976	0.976	0.976	0.976	0.964
	VGG	1	1	1	1	1	1
NB	IV3	0.994	0.917	0.915	0.922	0.917	0.879
	SN	0.995	0.976	0.976	0.978	0.976	0.965
	VGG	1	1	1	1	1	1

There is a constant superiority of 10-fold over 5-fold cross validation splitting schemes for all the classification accuracy metrics, i.e. for a total of 18 different combinations of three architectures and six classifiers. The highest difference between the two splitting schemes is observed for the MCC metrics with an average of 1% higher. A distinct higher classification accuracy is observed for SqueezeNet

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

architecture for most of the accuracy metrics and the highest values are observed when SqueezeNet is paired with Random Forest and AdaBoost classifiers. AUC, CA and F1 scores are generally high with small variations among all the architecture-classifier pairs. Precision exhibits some occasional dips from generally high values, and this could be attributed to the prioritization of recall over precision by some models. The relative dependence of recall over the precision dependence on the chosen architecture has been reported earlier in the literature [74] and our results here indicate a similar pattern. MCC shows a higher variation among the used metrics and this is an example of a metric that is more sensitive to the choice of the feature extractor. Neural network is observed to be the best classifier across all the feature extraction architectures used, indicating as an appropriate choice for this dataset and possibly for other datasets too.

4.2. Dataset 2 Results

In the Table 4 below are reported the classification performance based on six evaluation metrics as all features extracted from the three architectures were used when D2 is used instead of D1. Here the size of the images is 256 x 256 pixels and the total number of the images that are used is 3156 images with 66% of the images used for training and 34% used for testing. In terms, the average of all the metrics, VGG16 performs better (0.8969) than Inception V3 (0.8675) and SqueezeNet (0.8618) architectures and this could be because of its deeper architecture that can facilitate identification of key features that can provide higher classification accuracy. Its superiority is obvious particularly in the testing phase. Comparing the classifiers, a superiority of Neural Networks is observed (average of 0.9421 across all metrics) and this could be due to its characteristic of identifying complex patterns. Neural Network classifier is followed by KNN (0.9148), Random Forest (0.8936), AdaBoost (0.9107), Naïve Bayes (0.8326) and AdaBoost (0.8260). The observation on the underperformance of Decision Tree classifiers is a confirmation of prior results that are related with its difficulty in high dimensional data [75]. VGG16 architecture performs best when Neural Networks, KNN, and Random Forest classifiers are used with the cell images dataset. The metrics with higher reported values are AUC, CA and F1 scores. Inception V3 performs well when Neural Networks and KNN are used as classifiers. CA and REC for Inception V3 results are slightly lower than VGG16. SqueezeNet underperforms when compared to the other two architectures, particularly with Naïve Bayes and Decision Trees. Classifiers compare differently when combined with different architectures across different performance evaluation metrics. In terms of AUC, CA and F1 score, Neural Networks paired with the VGG16 constitute a good combination. Considering that KNN is a very simple classifier it still performs well when feature vectors extracted from VGG16 are used for classification. AdaBoost and Naïve Bayes perform poorly especially when combined with SqueezeNet.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

Table 4.
Comparative performance of six classifiers using three feature extraction architectures for three cell type classification for D2.

Training Process													
Classifier /Model		**AUC**		**CA**		**F1**		**PREC**		**REC**		**MCC**	
		5F	10F	5F	10F	5F	10F	5F	10F	5F	10F	5F	10F
KNN	IV3	0.99	0.991	0.943	0.946	0.942	0.946	0.943	0.947	0.943	0.946	0.914	0.92
	SN	0.982	0.981	0.916	0.92	0.915	0.92	0.915	0.92	0.916	0.92	0.874	0.881
	VGG	0.989	0.99	0.947	0.948	0.947	0.948	0.948	0.948	0.947	0.948	0.921	0.922
DT	IV3	0.899	0.906	0.879	0.89	0.879	0.89	0.879	0.89	0.879	0.89	0.818	0.835
	SN	0.875	0.878	0.863	0.861	0.863	0.861	0.863	0.861	0.863	0.861	0.795	0.792
	VGG	0.91	0.921	0.905	0.907	0.904	0.907	0.904	0.907	0.905	0.907	0.857	0.861
RF	IV3	0.982	0.983	0.914	0.92	0.915	0.92	0.915	0.921	0.914	0.92	0.872	0.88
	SN	0.976	0.979	0.902	0.908	0.902	0.909	0.903	0.909	0.902	0.908	0.853	0.863
	VGG	0.987	0.987	0.934	0.932	0.934	0.932	0.935	0.933	0.934	0.932	0.902	0.899
AB	IV3	0.897	0.901	0.862	0.869	0.862	0.868	0.862	0.868	0.862	0.869	0.793	0.803
	SN	0.874	0.879	0.831	0.838	0.831	0.839	0.831	0.839	0.831	0.838	0.747	0.758
	VGG	0.927	0.923	0.903	0.897	0.903	0.897	0.903	0.897	0.903	0.897	0.855	0.846
NN	IV3	0.998	0.997	0.974	0.978	0.974	0.978	0.974	0.978	0.974	0.978	0.962	0.967
	SN	0.996	0.996	0.964	0.965	0.964	0.965	0.964	0.965	0.964	0.965	0.945	0.948
	VGG	0.998	0.998	0.971	0.971	0.971	0.971	0.971	0.971	0.971	0.971	0.957	0.957
NB	IV3	0.959	0.959	0.876	0.876	0.876	0.876	0.877	0.877	0.876	0.876	0.814	0.814
	SN	0.918	0.918	0.797	0.795	0.795	0.794	0.796	0.794	0.797	0.795	0.696	0.694
	VGG	-	-	0.887	0.886	0.886	0.885	0.886	0.885	0.887	0.886	0.83	0.829
Testing Process													
Classifier /Model		AUC		CA		F1		PREC		REC		MCC	
KNN	IV3	0.936		0.831		0.833		0.854		0.831		0.756	
	SN	0.954		0.867		0.868		0.876		0.867		0.803	
	VGG	0.957		0.878		0.879		0.882		0.878		0.817	
DT	IV3	0.808		0.708		0.714		0.775		0.708		0.584	
	SN	0.825		0.794		0.796		0.81		0.794		0.697	
	VGG	0.856		0.833		0.836		0.846		0.833		0.753	
RF	IV3	0.923		0.751		0.758		0.808		0.751		0.647	
	SN	0.96		0.847		0.85		0.865		0.847		0.777	
	VGG	0.965		0.87		0.871		0.877		0.87		0.807	
AB	IV3	0.789		0.719		0.726		0.763		0.719		0.591	
	SN	0.818		0.758		0.76		0.778		0.758		0.643	
	VGG	0.849		0.799		0.804		0.824		0.799		0.706	
NN	IV3	0.981		0.867		0.87		0.89		0.867		0.809	
	SN	0.975		0.843		0.847		0.878		0.843		0.778	
	VGG	0.981		0.895		0.896		0.903		0.895		0.844	
NB	IV3	0.888		0.748		0.753		0.786		0.748		0.636	
	SN	0.936		0.821		0.819		0.827		0.821		0.736	
	VGG	-		0.818		0.819		0.82		0.818		0.727	

The reported results in Table 5 show the classification accuracies reported by six different metrics for two different splitting schemes and testing phase, 18 combinations of 3 feature extractor architectures with six classifiers. Here we use only 10 features that are extracted using Information Gain.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

Table 5.
Comparative performance of six classifiers using 10 relevant features obtained using dimensionality reduction from Information Gain for D2.

Training process													
Classifier /Model		**AUC**		**CA**		**F1**		**PREC**		**REC**		**MCC**	
		5F	**10F**	**5F**	**10F**	**5F**	**10F**	**5F**	**10F**	**5F**	**10F**	**5F**	**10F**
KNN	IV3	0.953	0.953	0.868	0.864	0.868	0.864	0.868	0.865	0.868	0.864	0.802	0.797
	SN	0.918	0.918	0.799	0.802	0.798	0.801	0.798	0.801	0.799	0.802	0.699	0.703
	VGG	0.953	0.953	0.872	0.875	0.872	0.875	0.872	0.875	0.872	0.875	0.808	0.813
DT	IV3	0.848	0.844	0.842	0.835	0.841	0.834	0.841	0.833	0.842	0.835	0.763	0.752
	SN	0.788	0.802	0.753	0.757	0.752	0.757	0.752	0.757	0.753	0.757	0.629	0.635
	VGG	0.864	0.877	0.838	0.851	0.838	0.851	0.837	0.85	0.838	0.851	0.758	0.777
RF	IV3	0.963	0.963	0.872	0.872	0.872	0.872	0.872	0.872	0.872	0.872	0.808	0.808
	SN	0.929	0.923	0.809	0.804	0.809	0.804	0.809	0.804	0.809	0.804	0.714	0.705
	VGG	0.962	0.963	0.879	0.883	0.879	0.883	0.879	0.883	0.879	0.883	0.819	0.824
AB	IV3	0.873	0.875	0.831	0.834	0.83	0.833	0.83	0.833	0.831	0.834	0.746	0.751
	SN	0.817	0.808	0.756	0.744	0.756	0.744	0.756	0.744	0.756	0.744	0.634	0.616
	VGG	0.876	0.879	0.834	0.838	0.834	0.838	0.834	0.839	0.834	0.838	0.752	0.758
NN	IV3	0.974	0.974	0.884	0.884	0.884	0.884	0.884	0.884	0.884	0.884	0.826	0.826
	SN	0.947	0.948	0.833	0.836	0.833	0.836	0.833	0.836	0.833	0.836	0.75	0.754
	VGG	0.975	0.975	0.898	0.901	0.897	0.9	0.898	0.901	0.898	0.901	0.847	0.851
NB	IV3	0.958	0.958	0.855	0.856	0.856	0.856	0.857	0.857	0.855	0.856	0.783	0.783
	SN	0.919	0.919	0.786	0.785	0.786	0.784	0.786	0.785	0.786	0.785	0.681	0.678
	VGG	0.946	0.946	0.829	0.831	0.828	0.83	0.828	0.831	0.829	0.831	0.745	0.748

Testing process							
Classifier /Model		AUC	CA	F1	PREC	REC	MCC
KNN	IV3	0.876	0.741	0.749	0.79	0.741	0.627
	SN	0.878	0.741	0.742	0.753	0.741	0.617
	VGG	0.931	0.834	0.835	0.837	0.834	0.751
DT	IV3	0.765	0.71	0.719	0.762	0.71	0.579
	SN	0.741	0.687	0.686	0.7	0.687	0.537
	VGG	0.852	0.818	0.819	0.821	0.818	0.727
RF	IV3	0.903	0.741	0.749	0.791	0.741	0.627
	SN	0.891	0.737	0.738	0.76	0.737	0.616
	VGG	0.943	0.838	0.839	0.843	0.838	0.757
AB	IV3	0.778	0.704	0.713	0.754	0.704	0.627
	SN	0.77	0.693	0.695	0.706	0.693	0.544
	VGG	0.86	0.813	0.815	0.818	0.813	0.72
NN	IV3	0.926	0.769	0.776	0.817	0.769	0.67
	SN	0.921	0.77	0.77	0.788	0.77	0.664
	VGG	0.946	0.839	0.841	0.847	0.839	0.76
NB	IV3	0.91	0.755	0.761	0.777	0.755	0.637
	SN	0.902	0.752	0.753	0.764	0.752	0.632
	VGG	0.942	0.83	0.829	0.83	0.83	0.745

When a limited number of features are used for classification, a decrease in the accuracy is observed as reported by all the metrics and this is expected. There is no significant difference in the training between the 5-fold and the 10-fold splitting schemes with the latter exhibiting slightly higher accuracy for all six metrics thus indicating a more favourable setting. But the 5-fold splitting scheme exhibits a smaller variation across all 18 combinations of architectures and classifiers and also a lower CoV (0.2-0.3) thus indicating a more favourable scheme in terms of robustness. The training accuracies vary from 97.5% (AUC), to ~90% (CA, F1 score, PREC and REC), and to the lowest value of 85.1% (MCC). The accuracy deteriorates by 7% to 19% when classification based on the use of all features compared to the

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

use of only 10 features. The largest decrease is observed for the MCC and the lowest is observed for the AUC metric. VGG16 achieves a higher accuracy when compared to the other two architectures for all the six classifiers used. The highest classification accuracies are found when VGG16 is paired with Neural Network classifier (AUC=94.60%), with Random Forest (AUC=94.3%) and with Naïve Bayes (AUC=94.2%). The average evaluation metrics for 18 architecture-classifier pairs for the testing phase are: AUC=87.42%, CA=76.51%, F1=76.83%, Prec=78.66%, REC=76.51% and MCC=65.76%.

The order of the evaluation metrics for the testing phase are ranked from the highest to the lowest as AUC (95.6%), PREC (84.7%), F1 score (84.1%), RECl and CA (83.9%) and MCC (76%).

5. Discussion

To compare the accuracy metrics results and to evaluate their robustness, coefficient of variation (CoV) is also reported as a percentage of the ratio of standard deviation over the mean of the metrics (6). This is a coefficient that is reported in the literature as an important metric to quantitatively compare different experiments [76, 77].

$$CoV = (\sigma/\mu) * 100 \quad (6)$$

The analysis is performed in Table 4, due to the high amount of training and testing sets present in D2 compared to D1. All features extracted by the architectures are used in order to have a better view of the architecture-classifier relation. CoV is minimal for each classifier when that classifier is paired with VGG16, and the lowest CoV is when Neural Network is paired with VGG16 (4.54%). VGG16 has the lowest CoV (4.44%), followed by Inception V3 (6.04%), and then SqueezeNet architecture (6.35%). When evaluating the performance of the testing of the classification for all the six accuracy evaluation metrics for 18 architecture-classifier pairs, generally the same patterns are observed. The highest testing accuracy is observed for the AUC metrics corresponding to VGG16 - Neural Network pair (AUC=98.10%) and the lowest testing accuracy is observed for the Inception V3 - AdaBoost pair (AUC=78.90%). For each classifier, its testing accuracy is highest when combined with VGG16 followed by its combination with SqueezeNet and it is lowest when combined with Inception V3, except when paired with Neural Network. The order of the testing accuracy for Neural Network as a classifier is in the decreasing order from 90.23% when combined with VGG16, 88.07% when combined with Inception V3 and 86.07% when combined with SqueezeNet. The careful analysis of the testing accuracy is very important as this really indicates whether the trained model can generalize on unseen data and whether it is useful in practical deployment (See Figure 7).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

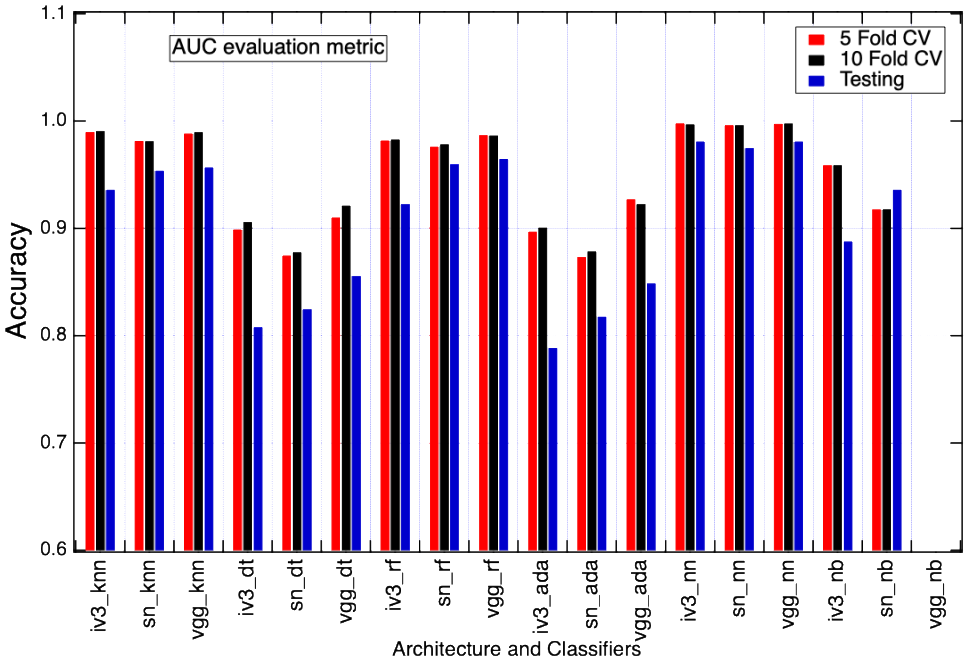

Figure 7.
AUC metric for 5-fold, 10-fold and testing for all the combinations of architecture and classifiers.

The testing accuracy as evaluated by six metrics exhibits its lowest CoV (4.47%) when Decision Tree classifier is used and this indicates the most robust classifier for different splitting schemes (5 fold and 10 fold), and also for testing over all metrics. The next lowest (best) CoV (4.54%) corresponds to the VGG16 architecture with Neural Network classifier. The constant superiority of one filter over the other for all six classifiers while using IV3 was an interesting observation that needs further analysis. This indicates a dependence of the classification on the cut off frequency of the filter that would be related to the features that can be preserved when a higher cut off frequency is used.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

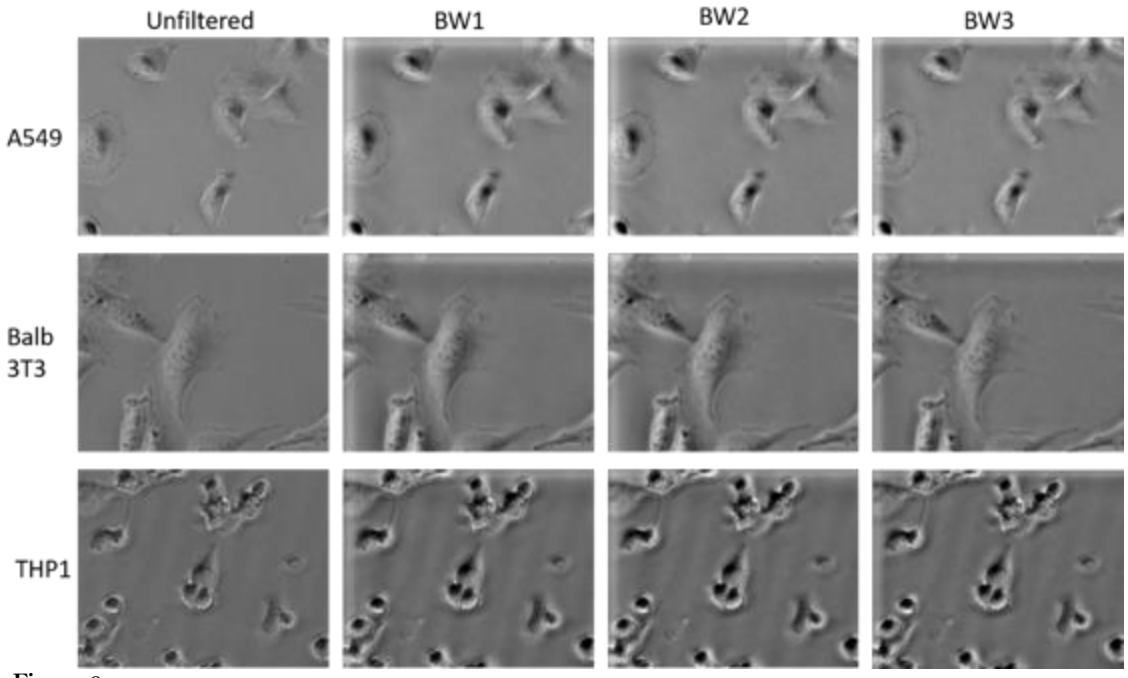

Figure 8.
Representation of the effect of the Butterworth filters on three cell types.

Subsequently, Butterworth filters (BW1, BW2, BW3) with different parameters (30/120/4, 30/150/4, 50/150/4) were applied to the image dataset, and the filtered images were used for training following the same procedure as for the unfiltered images (See Figure 8). The maximum accuracy metric (Max), its average, standard deviation (Stdev) and CoV are shown in Table 6 below (shown for AUC) corresponding to all 18 different combinations of architectures and classifiers.

Table 6.
Comparison of statistical measurements for AUC of 18 different classifier/architecture combination for filtered and unfiltered data.

	BW1	**BW2**	**BW3**	**Unfiltered**
Max.	98.00	97.80	97.20	98.10
Average	90.91	91.16	90.89	90.45
St. dev.	6.56	6.25	6.55	6.58
CoV.	7.21	6.85	7.20	7.28

A higher accuracy is observed for images that are filtered with all three filters when compared to the unfiltered dataset. BW2 exhibits a smaller variance across all 18 combinations and therefore a smaller CoV.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

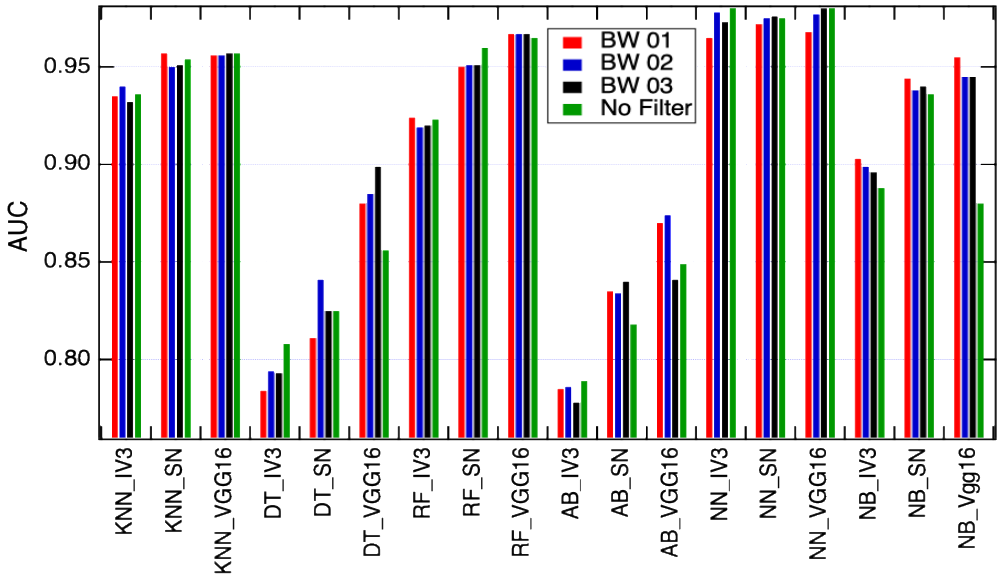

Figure 9.
AUC metric for three Butterworth filter and unfiltered dataset for all architecture-classifier pairs.

Significant differences in accuracy were observed among the filtered images, when comparing the results of 18 different classifier-architecture combination, given in Table 7. A graphical representation is also shown in Figure 9. Specifically, the accuracy obtained with filter BW1 was significantly higher than with filter BW3 ($p = 0.04$) when Decision Tree classifiers were combined with all three architectures. Furthermore, when the Decision Tree classifier was paired with VGG16, the accuracies for all three filters exceeded those of the unfiltered dataset. This indicates not only a significant difference between the same filter class but also a complete superiority of this filter class when compared to the unfiltered dataset.

Table 7.
Statistical measurements of 18 different classifier/architecture combination.

	BW1	BW2	BW3	Unfiltered
KNN_IV3	0.932	0.94	0.935	0.936
KNN_SN	0.951	0.95	0.957	0.954
KNN_VGG16	0.957	0.956	0.956	0.957
DT_IV3	0.793	0.794	0.784	0.808
DT_SN	0.825	0.841	0.811	0.825
DT_VGG16	0.899	0.885	0.88	0.856
RF_IV3	0.92	0.919	0.924	0.923
RF_SN	0.951	0.951	0.95	0.96
RF_VGG16	0.967	0.967	0.967	0.965
AB_IV3	0.778	0.786	0.785	0.789
AB_SN	0.84	0.834	0.835	0.818
AB_VGG16	0.841	0.874	0.87	0.849
NN_IV3	0.973	0.978	0.965	0.981
NN_SN	0.976	0.975	0.972	0.975
NN_VGG16	0.98	0.977	0.968	0.981
NB_IV3	0.896	0.899	0.903	0.888
NB_SN	0.94	0.938	0.944	0.936
NB_VGG16	0.945	0.945	0.955	0.88

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570

A significant difference is observed for the accuracies when BW1 and BW2 is compared (p=0.045) when IV3 is used with all six classifiers. No statistically significant differences were observed between the three Butterworth filters and the unfiltered dataset when SqueezeNet and VGG16 architectures are used.

6. Conclusions

Cell imaging is inherently a challenging problem, and this becomes more obvious in the case of microfluidics settings. One of the important experiments is both intraclass and interclass cell image classification. Here we systematically investigated the most optimal combination of architecture (among Inception V3, SqueezeNet, and VGG16) and classifier (from KNN, Decision Tree, Random Forest, AdaBoost, Neural Network, and Naïve Bayes) that can be used for robust supervised classification of different cells from unstained brightfield images. We observe a common superiority of the VGG16 architecture over the other two in the classification of the cell images and this indicates the benefit of a deeper architecture that would allow the extraction of more complex features from challenging texture and morphological structures. Among the classifiers, the Neural Network consistently achieved the highest classification metrics when paired with VGG16, demonstrating its ability to effectively utilize the rich feature representations extracted by the architecture. While other classifiers such as Random Forest and KNN performed well in specific scenarios, they were unable to match the flexibility and accuracy of Neural Networks across the different experimental conditions.

The use of Butterworth filters with several parameters as a preprocessing step proved to be useful in improving the classification accuracy and in indicating the importance of the parameters in differentiating the filters of the same family. The significant improvement in the accuracy as a function of the filter parameters indicates the importance of the cut off values and the order of the filter. Optimization of these two parameters critically balances the preservation of the features and the noise reduction. This underscores the importance of optimizing preprocessing techniques to the specific characteristics of the dataset to maximize model performance.

The results presented here are important for microscopy image analysis and particularly cell imaging. Fields including biomaterial risk assessment and drug delivery are distinct areas that require this type of analysis. The ability to accurately classify cellular responses to external stimuli, such as nano- and microstructured surfaces, is critical for evaluating cytotoxicity and understanding the interactions between cells and engineered materials. Future work could extend by evaluating modification in parameters of the architectures (for example pooling functions etc) and classifiers, including the analysis of broader set of filter parameters.

Funding:

This project received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 760921 (PANBioRA).

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

References

[1] B. J. Druker *et al.*, "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia," *New England Journal of Medicine*, vol. 344, no. 14, pp. 1031-1037, 2001. https://doi.org/10.1056/NEJM200104053441401

[2] G. Topalli, C. Xie, Y. Fan, L. Luan, R. Yin, and T. Chi, "A complementary pseudo-resistor with leakage current self-compensation for biopotential amplifiers.," presented at the 2023 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE. , 2023.

[3] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, no. 6096, pp. 816-821, 2012. https://doi.org/10.1126/science.1225829

[4] N. E. Vrana, *Introduction to biomaterials for tissue/organ regeneration. In Biomaterials for Organ and Tissue Regeneration*. Woodhead Publishing. https://doi.org/10.1016/B978-0-08-102906-0.00001-2, 2020.

[5] A. N. Halili, N. Hasirci, and V. Hasirci, "A multilayer tissue engineered meniscus substitute," *Journal of Materials Science: Materials in Medicine*, vol. 25, pp. 1195-1209, 2014. https://doi.org/10.1007/s10856-014-5164-5

[6] P. Sharma and J. P. Allison, "The future of immune checkpoint therapy," *Science*, vol. 348, no. 6230, pp. 56-61, 2015. https://doi.org/10.1126/science.aaa8172

[7] P. Baldi, "Deep learning in biomedical data science," *Annual Review of Biomedical Data Science*, vol. 1, no. 1, pp. 181-205, 2018. https://doi.org/10.1146/annurev-biodatasci-080917-013343

[8] X. Polisi, A. Halili, C. E. Tanase, A. Uka, N. E. Vrana, and A. Ghaemmaghami, *Computer assisted analysis of the hepatic spheroid formation. In Computational Bioengineering and Bioinformatics: Computer Modelling in Bioengineering*. Springer International Publishing. https://doi.org/10.1007/978-3-030-40115-5_10, 2020.

[9] S. Wu, X. Liu, K. W. Yeung, C. Liu, and X. Yang, "Biomimetic porous scaffolds for bone tissue engineering," *Materials Science and Engineering: R: Reports*, vol. 80, pp. 1-36, 2014. https://doi.org/10.1016/j.mser.2014.04.001

[10] Q. L. Loh and C. Choong, "Three-dimensional scaffolds for tissue engineering applications: Role of porosity and pore size," *Tissue Engineering Part B: Reviews*, vol. 19, no. 6, pp. 485–502, 2013. https://doi.org/10.1089/ten.teb.2012.0437

[11] S. A. Sell, P. S. Wolfe, K. Garg, J. M. McCool, I. A. Rodriguez, and G. L. Bowlin, "The use of natural polymers in tissue engineering: A focus on electrospun extracellular matrix analogues," *Polymers*, vol. 2, no. 4, pp. 522-553, 2010. https://doi.org/10.3390/polym2040522

[12] A. N. Halili, V. Bashari, I. Kazani, A. Uka, G. Alushllari, and F. Skuka, "Evaluating deep learning for segmenting nanobeads on nanofibers in electron microscopy images," presented at the NanoBalkan Conference 2024, Tirana, Albania, 2024.

[13] N. E. Vrana, K. Palm, and P. Lavalle, "Personalization of medical device interfaces: decreasing implant-related complications by modular coatings and immunoprofiling," *Future Science OA*, vol. 6, no. 8, p. FSO607, 2020. https://doi.org/10.2144/fsoa-2020-0070

[14] H. Knopf-Marques *et al.*, "Multifunctional polymeric implant coatings based on gelatin, hyaluronic acid derivative and chain length-controlled poly (arginine)," *Materials Science and Engineering: C*, vol. 104, p. 109898, 2019. https://doi.org/10.1016/j.msec.2019.109898

[15] X. Polisi, E. Dollani, and A. Uka, "Classification of cell biomaterial interaction toxicity level using convolutional neural networks," presented at the Ninth International Conference on Computational Bioengineering (ICCB 2022), Lisbon, Portugal, 2022.

[16] A. Chmayssem *et al.*, "Integrated on chip cytotoxicity test with real-time image analysis and electrochemical sensors-based monitoring," *Organ-On-Chip (EUROoCS 2021)*, vol. 1, no. 1, 2021.

[17] A. Chmayssem *et al.*, "New microfluidic system for electrochemical impedance spectroscopy assessment of cell culture performance: design and development of new electrode material," *Biosensors*, vol. 12, no. 7, p. 452, 2022. https://doi.org/10.3390/bios12070452

[18] A. Uka, A. Ndreu Halili, X. Polisi, A. O. Topal, G. Imeraj, and N. E. Vrana, "Basis of image analysis for evaluating cell biomaterial interaction using brightfield microscopy," *Cells Tissues Organs*, vol. 210, no. 2, pp. 77–104, 2021. https://doi.org/10.1159/000515210

[19] Y. Shao, X. Lu, S. Konijnenberg, C. Zhao, Y. Cai, and H. P. Urbach, "Spatial coherence measurement and partially coherent diffractive imaging using self-referencing holography," *Optics Express*, vol. 26, no. 4, pp. 4479-4490, 2018. https://doi.org/10.1364/OE.26.004479

[20] U. Bal, "Dual tree complex wavelet transform based denoising of optical microscopy images," *Biomedical Optics Express*, vol. 3, no. 12, pp. 3231-3239, 2012. https://doi.org/10.1364/BOE.3.003231

[21] A. Uka, G. Imeraj, B. Qesaraku, and B. Shehu, *Fourier ptychography microscopy resolution improvement employing refocusing*. Biodevices. https://doi.org/10.5220/0010770000003125, 2022.

[22] A. Uka, A. Tare, X. Polisi, and I. Panci, "FASTER R-CNN for cell counting in low contrast microscopic images," presented at the 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), IEEE, 2020.

[23] A. Uka, X. Polisi, A. Halili, C. Dollinger, and N. E. Vrana, "Analysis of cell behavior on micropatterned surfaces by image processing algorithms," presented at the IEEE EUROCON 2017-17th International Conference on Smart Technologies, IEEE. , 2017.

[24] X. Polisi, A. N. Halili, A. Uka, and C. Ciulla, "Two-stage unsupervised classification of cell health," presented at the 2023 International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE, 2023.

[25] N. Meng, E. Y. Lam, K. K. Tsia, and H. K.-H. So, "Large-scale multi-class image-based cell classification with deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2091-2098, 2018. https://doi.org/10.1109/JBHI.2018.2871665

[26] X. Polisi Duro, V. Sharka, A. Uka, D. A. Karras, and A. N. Halili, "Deep learning-based detection, segmentation and average area estimation of cells," *Journal of Natural and Technical Sciences*, vol. 29, no. 1, 2024.

[27] N. Meng, H. K. H. So, and E. Y. Lam, "Computational single-cell classification using deep learning on bright-field and phase images," presented at the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), IEEE, 2017.

[28] X. Polisi, D. Avdiu, A. Uka, A. N. Halili, K. Kollcaku, and C. Ciulla, "Evaluation of cell segmentation using pruning and quantization," presented at the 2023 International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE. , 2023.

[29] Z. Zheng, B. Sun, S. He, G. Wang, C. Bi, and T. Chen, "Unsupervised deep learning of bright-field images for apoptotic cell classification," *Signal, Image and Video Processing*, vol. 17, no. 7, pp. 3657-3664, 2023. https://doi.org/10.1007/s11760-023-02592-1

[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826, doi: https://doi.org/10.1109/CVPR.2016.308.

[31] R. Tamilarasi and S. Gopinathan, "Inception architecture for brain image classification," *Journal of Physics: Conference Series, IOP Publishing*, vol. 1964, no. 7, p. 072022, 2021. https://doi.org/10.1088/1742-6596/1964/7/072022

[32] C. Wang *et al.*, "Pulmonary image classification based on inception-v3 transfer learning model," *IEEE Access*, vol. 7, pp. 146533-146541, 2019. https://doi.org/10.1109/ACCESS.2019.2946000

[33] F. N. Iandola, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.

[34] M. Tsivgoulis, T. Papastergiou, and V. Megalooikonomou, "An improved SqueezeNet model for the diagnosis of lung cancer in CT scans," *Machine Learning with Applications*, vol. 10, p. 100399, 2022. https://doi.org/10.1016/j.mlwa.2022.100399

[35] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. Huang, "An improved VGG16 model for pneumonia image classification," *Applied Sciences*, vol. 11, no. 23, p. 11185, 2021. https://doi.org/10.3390/app112311185

[36] R. S. Chugh, V. Bhatia, K. Khanna, and V. Bhatia, "A comparative analysis of classifiers for image classification," presented at the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2020.

[37] E. O. Omuya, G. O. Okeyo, and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," *Expert Systems with Applications*, vol. 174, p. 114765, 2021. https://doi.org/10.1016/j.eswa.2021.114765

[38] S. Raghavendra and M. Indiramma, "Hybrid data mining model for the classification and prediction of medical datasets," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 5, no. 3-4, pp. 262-284, 2016. https://doi.org/10.1504/IJKESDP.2016.10001566

[39] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data classification: Algorithms and applications*, p. 37, 2014. https://doi.org/10.1201/b17320-3

[40] R. Mehta, C. Shui, B. Nichyporuk, and T. Arbel, "Information gain sampling for active learning in medical image classification. In International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging," presented at the Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-16749-2_13, 2022.

[41] A. Uka, X. Polisi, J. Barthes, A. N. Halili, F. Skuka, and N. E. Vrana, "Effect of preprocessing on performance of neural networks for microscopy image classification," presented at the 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE, 2020.

[42] S. Minoshima *et al.*, "Optimization of Butterworth filter for brain SPECT imaging," *Annals of Nuclear Medicine*, vol. 7, pp. 71-77, 1993. https://doi.org/10.1007/bf03164571

[43] M. Lyra and A. Ploussi, "Filtering in SPECT image reconstruction," *International Journal of Biomedical Imaging*, vol. 2011, no. 1, p. 693795, 2011. https://doi.org/10.1155/2011/693795

[44] I. Mendieta, M. Rodríguez-Nieto, R. E. Nuñez-Anita, J. L. Menchaca-Arredondo, G. García-Alcocer, and L. C. Berumen, "Ultrastructural changes associated to the neuroendocrine transdifferentiation of the lung adenocarcinoma cell line A549," *Acta Histochemica*, vol. 123, no. 8, p. 151797, 2021. https://doi.org/10.1016/j.acthis.2021.151797

[45] I. Lasocka *et al.*, "Cytocompatibility of graphene monolayer and its impact on focal cell adhesion, mitochondrial morphology and activity in BALB/3T3 fibroblasts," *Materials*, vol. 14, no. 3, p. 643, 2021. https://doi.org/10.3390/ma14030643

[46] C. Dollinger *et al.*, "Controlling incoming macrophages to implants: responsiveness of macrophages to gelatin micropatterns under M1/M2 phenotype defining biochemical stimulations," *Advanced Biosystems*, vol. 1, no. 6, p. 1700041, 2017. https://doi.org/10.1002/adbi.201700041

[47] A. Dogra and P. Bhalla, "Image sharpening by gaussian and butterworth high pass filter," *Biomedical and Pharmacology Journal*, vol. 7, no. 2, pp. 707-713, 2014. https://doi.org/10.13005/bpj/545

[48] Y. P. Wang, Q. Wu, and K. R. Castleman, *Image enhancement. In Microscope Image Processing*. Academic Press. https://doi.org/10.1016/B978-0-12-821048-9.00004-5, 2023.

[49] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence*, vol. 52, no. 4, pp. 4543-4581, 2022. https://doi.org/10.1007/s10489-021-02680-3

[50] Q. Guan *et al.*, "Deep convolutional neural network Inception-v3 model for differential diagnosing of lymph node in cytological images: a pilot study," *Annals of Translational Medicine*, vol. 7, no. 14, pp. 1-9, 2019. https://doi.org/10.21037/atm.2019.06.29

[51] N. Dong, L. Zhao, C.-H. Wu, and J.-F. Chang, "Inception v3 based cervical cell classification combined with artificially extracted features," *Applied Soft Computing*, vol. 93, p. 106311, 2020. https://doi.org/10.1016/j.asoc.2020.106311

[52] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of Digital Imaging*, vol. 32, pp. 605-617, 2019. https://doi.org/10.1007/s10278-019-00182-9

[53] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha, "Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images," *EBioMedicine*, vol. 27, pp. 317-328, 2018. https://doi.org/10.1016/j.ebiom.2017.12.026

[54] B. Kieffer, M. Babaie, S. Kalra, and H. R. Tizhoosh, "Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks," presented at the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), 2017.

[55] C. A. Ferreira *et al.*, "Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2," presented at the International Conference Image Analysis and Recognition, Springer International Publishing, 2018.

[56] D. Arifianto and A. S. Agoes, "Cervical cancer image classification using CNN transfer learning," presented at the 2nd International Seminar of Science and Applied Technology (ISSAT 2021), Atlantis Press, 2021.

[57] K. Nakamichi, H. Lu, H. Kim, K. Yoneda, and F. Tanaka, "Classification of circulating tumor cells in fluorescence microscopy images based on SqueezeNet," presented at the 2019 19th International Conference on Control, Automation and Systems (ICCAS), IEEE, 2019.

[58] E. Başaran, "Classification of white blood cells with SVM by selecting SqueezeNet and LIME properties by mRMR method," *Signal, Image and Video Processing*, vol. 16, no. 7, pp. 1821-1829, 2022. https://doi.org/10.1007/s11760-022-02141-2

[59] A. S. Agoes, Z. Hu, and N. Matsunaga, "Fine tuning based SqueezeNet for vehicle classification," in *Proceedings of the International Conference on Advances in Image Processing*, 2017, pp. 14-18, doi: https://doi.org/10.1145/3133264.3133270.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[61] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," *Applied Intelligence*, vol. 51, no. 5, pp. 2850-2863, 2021. https://doi.org/10.1007/s10489-020-02055-x

[62] E. Gavas and K. Olpadkar, "Deep CNNs for peripheral blood cell classification," *arXiv preprint arXiv:2110.09508*, 2021. https://doi.org/10.48550/arXiv.2110.09508

[63] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. A. Almansour, "Deep CNN model based on VGG16 for breast cancer classification," presented at the 2021 International Conference on Information Technology (ICIT), IEEE. , 2021.

[64] T. Kaur and T. K. Gandhi, "Automated brain image classification based on VGG-16 and transfer learning," presented at the 2019 International Conference on Information Technology (ICIT), IEEE, 2019.

[65] R. B. Pereira, A. P. D. Carvalho, B. Zadrozny, and L. H. C. Merschmann, "Information gain feature selection for multi-label classification," *Journal of Information and Data Management*, vol. 6, no. 1, 2015. https://doi.org/10.5753/jidm.2015.294

[66] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, Springer Berlin Heidelberg*, 2003, pp. 986-996, doi: https://doi.org/10.1007/978-3-540-39964-3_62.

[67] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015. https://doi.org/10.11919/j.issn.1002-0829.215044

[68] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31-39, 2017. https://doi.org/10.17849/insm-47-1-31-39.1

[69] R. E. Schapire, *Explaining adaboost. In Empirical inference: Festschrift in honor of Vladimir N. Vapnik*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_5, 2013.

[70] K. Gurney, *An introduction to neural networks*. CRC Press. https://doi.org/10.1201/9781315271646, 2018.

[71] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia*, vol. 18, no. 60, pp. 1-8, 2006. https://doi.org/10.1007/978-0-387-30164-8_5

[72] V. K. Verma, K. Saxena, and U. Banodha, *Analysis effect of k values used in K Fold cross validation for enhancing performance of machine learning model with decision tree. In: Garg, D., Rodrigues, J.J.P.C., Gupta, S.K., Cheng, X., Sarao, P., Patel, G.S. (eds) Advanced Computing. IACC 2023. Communications in Computer and Information Science*. Cham: Springer. https://doi.org/10.1007/978-3-031-56700-1_30, 2024.

[73] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of machine learning algorithms with different K values in K-fold CrossValidation," *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61-71, 2021. https://doi.org/10.5815/ijitcs.2021.06.05

[74] N. Nasir, M. Ahmed, N. Afreen, and M. Sameer, "Alzheimer's magnetic resonance imaging classification using deep and meta-learning models," *arXiv preprint arXiv:2405.12126*, 2024.

[75] W. Liu and I. W. Tsang, "Making decision trees feasible in ultrahigh feature and label dimensions," *Journal of Machine Learning Research*, vol. 18, no. 81, pp. 1-36, 2017.

[76] D. A. Powell, L. M. Anderson, R. Y. Cheng, and W. G. Alvord, "Robustness of the Chen–Dougherty–Bittner procedure against non-normality and heterogeneity in the coefficient of variation," *Journal of Biomedical Optics*, vol. 7, no. 4, pp. 650-660, 2002. https://doi.org/10.1117/1.1501561

[77] Y. Yu, A. R. Khan, and J. Xu, "Measuring robustness for NLP," in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 3908-3916.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 682-707, 2025
DOI: 10.55214/25768484.v9i2.4570