Single-Cell Conventional Pap Smear Image Classification Using Pre-Trained Deep Neural Network Architectures

Mohammed Aliy Mohammed (✉ mohaliyet@gmail.com)

Jimma University

Fetulhak Abdurahman

Jimma University

Yodit Abebe Ayalew

Hawassa University

Research Article

Keywords: deep learning, image classification, cervical cancer, Pap smear

Posted Date: March 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-257282/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Single-cell conventional Pap smear image classification using pre-trained deep neural network architectures

Corresponding Author: Mohammed Aliy Mohammed (Lecturer, School of Biomedical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia); Email: mohaliyet@gmail.com

Co-author: Fetulhak Abdurahman (Lecturer, Faculty of Electrical and Computer Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia); Email: afetulhak@yahoo.com

Co-author: Yodit Abebe Ayalew (Lecturer, Department of Biomedical Engineering, Hawassa Institute of Technology, Hawassa University, Hawassa, Ethiopia); Email: yoditabebe9391@gmail.com

Abstract

Background: Automating Pap smear based cervical cancer screening could alleviate the challenge related to the shortage of skilled manpower, mainly pathologists in developing countries. The astonishing accuracy and reproducibility of deep neural networks in computer vision have made them a frontline candidate to tackle such challenges. In this regard, the main purpose of this research project is to classify single-cell Pap smear microscopic images using the architectures of the pre-trained deep convolutional neural

network image classifiers. We have trained five class single-cell Pap smear images from SIPaKMeD on top ten pre-trained image classification architectures. The architectures were selected from Keras Applications based on their top 1% accuracy. Then, the output layers of the selected architectures were fine-tuned from 1000 classes to five classes and retrained.

Results: Our experimental result demonstrated that DenseNet169 outperformed the selected 10 pre-trained architectures with an average accuracy, precision, recall and f1-scores of 0.990, 0.974, 0.974 and 0.974, respectively. Moreover, it dashed the benchmark accuracy proposed by the creators of the dataset with 3.70%.

Conclusions: Even though the size of DenseNet169 is small compared to the experimented architectures, still it is not suitable for mobile and edged devices. It is required to experiment with mobile or small size image classification neural network architectures.

Keywords: deep learning, image classification, cervical cancer, Pap smear

Background

Cervical cancer is a women-specific sexually transmitted infectious disease caused by, mainly, high-risk human papillomavirus (HPV). Worldwide, an estimated 570,000 cases and 311,000 deaths were registered in 2018 only. Among these numbers, about 85% of them are from developing countries (1).

Considering the prevalence of the disease, international organizations such as the World health organization (WHO) start to set new initiatives to eliminate it from the public health burden. The WHO's new strategy emphasized on the elimination of cervical cancer from public health problems by 2030 mainly focusing on three pillars (prevention, screening and treatment/ management) in a comprehensive approach. In the strategy, it is clearly stated

that to reach the stage of cervical cancer elimination, every country must give 90% coverage of HPV vaccine for girls of 15 years of age, perform 70% high-performance cervical cancer test (screening) for females between ages of 35 and 45, treat 90% of precancerous lesions and 90% management of invasive cancer patients (2).

In the past few decades, high-income countries have implemented population-wide screening programs and showed a significant reduction in mortality and morbidity caused by cervical cancer (3,4). The experience could be a good model to be further extended in low- and middle-income countries. However, the lack of basic resources such as skilled health personnel and screening tools have been posing a major challenge (5,6).

The latest WHO guideline regarding cervical cancer screening recommends three main techniques: high-risk HPV type testing using polymerase chain reaction (PCR), visual inspection with acetic acid (VIA), and cervical cytology (7). Among the three, cervical cytology is the most common and the orthodox way of screening. It has been considered as the standard technique valuing its contribution to the reduction of incidence and mortality rate in many high-income countries worldwide (5). The popular and well-developed techniques for cervical cytology are conventional Papanicolaou smears (CPS) and liquid-based cytology (LBC). The results of comparative studies focusing on the quality of CPS and LBC samples concluded that LBC is better than CPS (8,9). However, considering the economic burden, LBC is more common in high-income countries whereas CPS is more preferable in low- and middle-income countries (8).

Even though cytology techniques are effective in the reduction of morbidity and mortality, they suffered from the following main drawbacks: their sensitivity is less optimal, the interpretation of the result mainly depends on the morphological characteristics of the

65 cytoplasm and nucleus of the cell which requires a highly skilled cytotechnologist.

66 Moreover, analyzing a single sample takes considerably longer time and labor-intensive.

67 To bridge the aforementioned gaps of cervical cytology screening, computer vision experts

68 have been proposing semi- or fully-automated computer-aided analysis tools, especially for

69 the CPS ones. They developed systems that either classify single-cell CPS images or detect

70 abnormal cells from full-slide CPS images. A detailed and extended review is found in (10).

71 In general, as illustrated in Figure 1, three single-cell CPS image analysis pipelines have been

72 proposed in the literature. The traditional techniques follow either pipeline 1 or pipeline 2

73 or both combined. The main difference between the two pipelines is the requirement of the

74 segmentation stage. For instance, if the required feature vectors are attributes of the

75 morphology or shape of an object such as area, perimeter, thinness ratio, and eccentricity,

76 first, the cytoplasm or the nucleus of the cells need to be segmented from the rest of the

77 image content. On the other hand, if the required features do not require descriptors of

78 segmented objects such as chromatin and texture, the segmentation stage will be skipped

79 as it is depicted in the pipeline 2. In other words, the feature vectors will be directly

80 calculated from the pre-processed CPS images. Features calculated using the two pipelines

81 commonly known as hand-crafted features. Hand-crafted features give a privilege to the

82 computer vision expert to select and supervise the extracted feature vectors. Sometimes

83 dimensionality reduction schemes pick the right subset from a bucket of large feature

84 vectors. Approaches that follow pipeline 1 and pipeline 2 have presented in (11–18). The

85 other technique (pipeline 3) takes the benefit of deep convolutional neural networks (CNNs)

86 to learn complex features directly from the labelled raw images. The main advantage of

87 these deep CNNs is their ability to extract feature vectors without the intervention of

computer vision experts. Works presented in (19–24) are good examples of pipeline 3. In this study, we followed pipeline 3 and leveraged pre-trained deep learning architectures considering their recent remarkable performance improvement for image classification tasks.

Figure 1. Common pipelines to classify CPS images: classification using hand-crafted features and classification using deep learning schemes.

Plissiti, M.E et al. (17) produced a new CPS dataset in 2018 named SIPaKMeD and benchmarks for the traditional and deep learning based classification computer vision projects. For the deep learning based classification benchmark, firstly, they used the architecture of VGG-19 and fine-tuned the last layer to make it suitable for a 5-classes. Secondly, they used the pre-activated CNN of the model as a feature extractor and employed a support vector machine (SVM) as a classifier. Thirdly, they extracted the features by activating the CNN of the model to extract the feature maps before classifying the images into 5 classes using SVM. They achieved an average accuracy of 95.35%, 93.35%

and 94%, respectively. Except for the benchmark, to the authors' knowledge of the researchers, there are no deep learning based works that base their work on the SIPaKMeD dataset.

In this study, our main aim was to classify single-cell CPS images from SIPaKMeD dataset into five classes using the fine-tuned architectures of pre-trained image classification deep neural networks.

Experiment and Results

To maintain a fair comparison, all the training hyperparameters were kept identical in all experimental setups. As illustrated in **Error! Reference source not found.**Figure 2 and Figure 3 the networks were trained over 100 epochs with a loss function of categorical cross-entropy, a batch size of 32 and adagrad optimizer. Also, we have trained all the models with an initial learning rate of 0.001 which changes its value by a factor of 0.5 if there is no increment in the validation accuracy over 10 consecutive epochs until it reaches a value of 0.00001.

Figure 2. . Training accuracy (left) and training loss (right) of the architectures over 100 epochs.

Figure 3. Validation accuracy (left) and validation loss (right) of the architectures over 100 epochs.

After the training, we evaluated the architectures using the test dataset and their evaluation results are summarized in Table 1.

Discussion

As it can be inferred from Table 1, DenseNet169 outperform the rest of the architectures in all evaluation metrics. The normalized average accuracy, precision, recall and f1-score values are 0.990, 0.974, 0.974 and 0.974, respectively. Almost in all experiments, Koilocytotic cells are the challenging one to classify, i.e. their true positive value is the least compared to the others. Similar reporting can be found in the benchmark manuscript (17). In the second place, metaplastic cells are to be classified incorrectly.

When we further inspected the aforementioned cell types, we found out that most of the false negatives of Koilocytotic cells were incorrectly classified as metaplastic and most of the metaplastic cells were incorrectly classified as Koilocytotic cells. That means, the experiment tells us the need to increase the data variation between the two classes. As an example, see

the confusion matrix of DenseNet169 in Figure 4. Five of Koilocytotic cells are classified as Metaplastic and the vice versa.

		Accuracy	Precision	Recall	F1-score			Accuracy	Precision	Recall	F1-score
NASNetLarge	DC	0.992	0.980	0.980	0.980	InceptionResNetV2	DC	0.992	0.962	1.000	0.980
	KC	0.964	0.894	0.930	0.912		KC	0.958	0.899	0.890	0.894
	MC	0.970	0.913	0.940	0.926		MC	0.968	0.904	0.940	0.922
	PC	0.994	0.990	0.980	0.985		PC	0.988	0.980	0.960	0.970
	SIC	0.988	1.000	0.940	0.969		SIC	0.990	1.000	0.950	0.974
	Average	0.982	0.955	0.954	0.954		Average	0.979	0.949	0.948	0.948
Xception	DC	0.990	0.961	0.990	0.975	ResNet152V2	DC	0.986	0.952	0.980	0.966
	KC	0.968	0.920	0.920	0.920		KC	0.958	0.891	0.900	0.896
	MC	0.974	0.922	0.950	0.936		MC	0.968	0.904	0.940	0.922
	PC	0.986	0.989	0.940	0.964		PC	0.992	0.990	0.970	0.980
	SIC	0.998	1.000	0.990	0.995		SIC	0.988	1.000	0.940	0.969
	Average	0.983	0.959	0.958	0.958		Average	0.978	0.947	0.946	0.946
InceptionV3	DC	0.988	0.943	1.000	0.971	DenseNet201	DC	0.988	0.961	0.980	0.970
	KC	0.964	0.936	0.880	0.907		KC	0.964	0.918	0.900	0.909
	MC	0.966	0.888	0.950	0.918		MC	0.978	0.941	0.950	0.945
	PC	0.984	0.979	0.940	0.959		PC	1.000	1.000	1.000	1.000
	SIC	0.994	1.000	0.970	0.985		SIC	0.998	1.000	0.990	0.995
	Average	0.979	0.949	0.948	0.948		Average	0.986	0.964	0.964	0.964
ResNet101V2	DC	0.986	0.951	0.980	0.966	ResNet152	DC	0.992	0.971	0.990	0.980
	KC	0.964	0.918	0.900	0.909		KC	0.968	0.912	0.930	0.921
	MC	0.962	0.893	0.920	0.906		MC	0.974	0.939	0.930	0.935
	PC	0.994	0.980	0.990	0.985		PC	0.996	1.000	0.990	0.995
	SIC	0.990	1.000	0.950	0.974		SIC	0.992	1.000	0.980	0.990
	Average	0.979	0.949	0.948	0.948		Average	0.986	0.964	0.964	0.964
ResNet101	DC	0.992	0.980	0.980	0.980	DenseNet169	DC	0.998	1.000	0.990	0.995
	KC	0.962	0.909	0.900	0.905		KC	0.974	0.922	0.950	0.936
	MC	0.972	0.913	0.950	0.931		MC	0.978	0.941	0.950	0.945
	PC	0.998	1.000	0.990	0.995		PC	0.998	1.000	0.990	0.995
	SIC	0.996	1.000	0.980	0.990		SIC	0.998	1.000	0.990	0.995
	Average	0.984	0.961	0.960	0.960		Average	0.990	0.974	0.974	0.974

Table 1. Individual and average accuracies, precisions, recalls and F1-scores of the deep neural architectures when they evaluated using the test dataset.

Figure 4. Confusion matrix for classification of the test dataset using fine-tuned DenseNet169.

Another aspect of the experiment we need to give an emphasis is the size of the weight files. DenseNet169 has the smallest weight size (Table 2 shows the size of the original weight file). However, the size is not suitable to deploy to mobile and edge devices. It is required to experiment with small size and mobile image classification architectures.

Finally, we compared our finding with the deep convolutional neural network benchmark results reported by the dataset creators. In their work, they presented an average accuracy of 95.35 ± 0.42 % using the fine-tuned VGG19 network as a feature extractor and softmax as

a classifier. In this research, we achieved a normalized average accuracy of 0.990 which is way better than the benchmark.

Conclusion

In this paper, we experimented with the top ten Keras Applications deep learning architectures selected based on their top-1 accuracy for image classification tasks. We used the selected architectures to analyze their performance to classify single-cell CPS images. All the architectures were re-trained using SIPaKMeD dataset by changing the output layer of the architectures from 1000 classes used in ImageNet to 5 classes of CPS images. From the selected 10 pre-trained architectures, DenseNet169 outperformed the other architectures and achieved state-of-the-art performance compared to the benchmark result generated by the SIPaKMeD dataset creators. Using DenseNet169 a normalized average accuracy of 0.990 was achieved which is greater than the benchmark by approximately 3.7%. In the future, further experimentation with small size and mobile architectures is required to make the size of model weights suitable for mobile and edge devices.

Materials and methods

Dataset

In this study, a recently introduced publicly available dataset named SIPaKMeD was used (17). The dataset contains a total number of 4049 single-cell images that were manually cropped-out from 966 full-slide Pap smear images. The cells were grouped based on their abnormality and benign level as - superficial-intermediate cells (SIC), parabasal cells (PC), koilocytotic cells (KC), dyskeratotic cells (DC) and metaplastic cells (MC). The first two are normal, the second two are abnormal and the last one is benign. The distribution of images

across the single-cell image classes is seemingly uniform - 831, 787, 825, 793 and 813, respectively. Figure 5 shows representative images of the five classes.

Figure 5. Sample images from the SIPaKMeD dataset: superficial-intermediate (a), parabasal (b), koilocytotic (c), metaplastic (d) and dyskeratotic (e) cells.

Data Preparation

The dataset was partitioned into train, validation and test sets. The test data was prepared by taking randomly 100 single cell images from each cell category which accounts for nearly 12% of the total dataset. The remaining dataset was partitioned into 80% for training and 20% for validation.

As depicted in the pipeline of this study (see Figure 6), during the data pre-processing stage all images were resized to 128x128 before being fed to the classification models to keep the pre-trained architecture requirements. Image normalization was done to keep the dynamic range of pixel intensities of the image between 0 and 1. During the model training, we applied affine transformation using horizontal and vertical flipping and rotation in the range of -45° and 45° to increase intraclass variation. Even though the cross-class distribution is considerably uniform (the ratio between the classes with the smallest to the largest number of images is approximately 0.95), at the end of the preprocessing stage, class weight balancing was applied to the training and validation partitions of the dataset using Equation

1. At the time of training, the distribution of the classes for individual batches turned out to be 0.97, 1.03, 0.98, 1.02 and 1.00 for SIC, PC, KC, MC and DC, respectively.

$$w_j = \frac{S}{n * s_j}$$

Where, w_j stands for the weight of the class j , S for the total number of samples, n for the number classes and s_j for the samples in the class j .

Figure 6. The general pipeline of the research project: image acquisition, pre-processing, feature extraction and classification.

Proposed Approach

In this study, as illustrated in Table 1, we selected 10 popular pre-trained deep learning architectures for image classification from Keras applications (25) based on their top-1 accuracy. Top-1 accuracy refers to the normalized performance of a model to predict exactly the expected answer. For example, the probability of NASNetLarge to predict exactly the first answer is 0.823 out of a unit scale. The selected modes were trained on ImageNet (26) - a dataset of 1000 classes of natural images.

Model	Size	Top-1 Accuracy
NASNetLarge	343 MB	0.825
InceptionResNetV2	215 MB	0.803
Xception	88 MB	0.790
ResNet152V2	232 MB	0.780
InceptionV3	92 MB	0.779
DenseNet201	80 MB	0.773
ResNet101V2	171 MB	0.772
ResNet152	232 MB	0.766
ResNet101	171 MB	0.764
DenseNet169	57 MB	0.762

Table 2. Proposed pre-trained classification models weight size and their top-1 accuracy performance on the ImageNet's validation dataset.

Considering the dataset size (enough for deep learning applications) and the image type (microscopic image – not natural image), the selected learning approach was re-training the entire architectures after fine-tuning the classifier layer. In other words, the architectures of the feature extraction base were re-trained again using the CSP dataset to populate it with new weights and the output layer was fine-tuned from 1000 classes down to 5 classes. To converge the output of the feature extraction base from 4D tensor to a 2D tensor an average pooling layer was introduced. In the end, fully connected links were created between the pooling layer and the output dense layer as indicated in Figure 6. All the experiments were performed using Google's free cloud platform Kaggle with Nvidia Tx1008 GPU and 12 GB of RAM.

Evaluation Metrics

We evaluated the performance of the classification models using four objective evaluation metrics including accuracy, precision, recall and f1-score. The metrics base their mathematical foundation on the true positive (TP), true negative (TN), false negative (FN) and false positive (FP) values of the

models' prediction. A comprehensive summary of the metrics is found in (27) and their mathematical formulation as follows.

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

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right)$$

List of abbreviations

CNN – Convolutional Neural Network
CPS – Conventional Papanicolaou Smears
DC – Dyskeratotic Cells
FN – False Negative
FP – False Positive
KC – Koilocytotic Cells
LBC – Liquid Based Cytology
MC – Metaplastic Cells
PC – Parabasal Cells
SIC – superficial-Intermediate Cells
TN – True Negative
TP – True Positive
SVM – Support Vector Machine
WHO – World Health Organization
VIA – Visual Inspection with Acetic acid

Acknowledgments

Not applicable.

Authors' contributions

MA, FA and YA contributed to the design of the study. MA experimented, analyzed the results and drafted the manuscript. FA and YA proof read and edited the manuscript. MA, FA and YA have approved the final version of the manuscript and agreed to be accountable for all aspects of the work. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The minimal dataset used for this study was extracted from the original SIPaKMeD dataset.

- Link to the original dataset - <https://www.cs.uoi.gr/~marina/sipakmed.html>
- Link to the minimal dataset - <https://www.kaggle.com/mohaliy2016/papsinglecell/>

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no any competing interest.

Author details

Mohammed Aliy Mohammed (Lecturer, School of Biomedical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia); Email: mohaliyet@gmail.com

265 Fetulhak Abdurahman (Lecturer, Faculty of Electrical and Computer Engineering, Jimma
 266 Institute of Technology, Jimma University, Jimma, Ethiopia); Email: afetulhak@yahoo.com
 267 Yodit Abebe Ayalew (Lecturer, Department of Biomedical Engineering, Hawassa Institute of
 268 Technology, Hawassa University, Hawassa, Ethiopia); Email: yoditabebe9391@gmail.com

269 References

- 270 1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al.
 271 Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and
 272 methods [Internet]. Vol. 144, International Journal of Cancer. Wiley-Liss Inc.; 2019
 273 [cited 2021 Feb 11]. p. 1941–53. Available from:
 274 <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31937>
- 275 2. WHO. Draft Global strategy towards eliminating cervical cancer as a public health
 276 problem. 2019; Available from: <https://bit.ly/2Ss79ue>
- 277 3. Autier P, Sullivan R. Population Screening for Cancer in High-Income Settings: Lessons
 278 for Low- and Middle-Income Economies. J Glob Oncol. 2019 Dec;(5):1–5.
- 279 4. Vale DB, Bragança JF, Zeferino LC. Cervical Cancer Screening in Low- and Middle-
 280 Income Countries. In: Uterine Cervical Cancer [Internet]. Cham: Springer International
 281 Publishing; 2019. p. 53–9. Available from: [http://link.springer.com/10.1007/978-3-](http://link.springer.com/10.1007/978-3-030-02701-8_4)
 282 [030-02701-8_4](http://link.springer.com/10.1007/978-3-030-02701-8_4)
- 283 5. Catarino R, Petignat P, Dongui G, Vassilakos P. Cervical cancer screening in developing
 284 countries at a crossroad: Emerging technologies and policy choices. Vol. 6, World
 285 Journal of Clinical Oncology. Baishideng Publishing Group Co., Limited; 2015. p. 281–
 286 90.

- 287 6. Beddoe AM. Elimination of cervical cancer: Challenges for developing countries.
288 Ecanermedicalscience. 2019 Nov 12;13.
- 289 7. World Health Organization. WHO | Guidelines for screening and treatment of
290 precancerous lesions for cervical cancer prevention. 2013 [cited 2020 Jun 1];
291 Available from:
292 http://www.who.int/reproductivehealth/publications/cancers/screening_and_treatm
293 [ent_of_precancerous_lesions/en/](http://www.who.int/reproductivehealth/publications/cancers/screening_and_treatment_of_precancerous_lesions/en/)
- 294 8. De Bekker-Grob EW, De Kok IMCM, Bulten J, Van Rosmalen J, Vedder JEM, Arbyn M,
295 et al. Liquid-based cervical cytology using ThinPrep technology: Weighing the pros
296 and cons in a cost-effectiveness analysis. Cancer Causes Control [Internet]. 2012 Aug
297 17 [cited 2020 Jun 10];23(8):1323–31. Available from:
298 <http://link.springer.com/10.1007/s10552-012-0011-1>
- 299 9. Haghighi F, Ghanbarzadeh N, Ataee M, Sharifzadeh G, Mojarad J, Najafi-Semnani F. A
300 comparison of liquid-based cytology with conventional Papanicolaou smears in
301 cervical dysplasia diagnosis. Adv Biomed Res. 2016;5(1):162.
- 302 10. Conceição T, Braga C, Rosado L, Vasconcelos MJM. A review of computational
303 methods for cervical cells segmentation and abnormality classification. Vol. 20,
304 International Journal of Molecular Sciences. MDPI AG; 2019.
- 305 11. Plissiti ME, Nikou C. Cervical cell classification based exclusively on nucleus features.
306 In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial
307 Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg; 2012.
308 p. 483–90.

- 309 12. Chen YF, Huang PC, Lin KC, Lin HH, Wang LE, Cheng CC, et al. Semi-automatic
310 segmentation and classification of pap smear cells. IEEE J Biomed Heal Informatics.
311 2014 Jan;18(1):94–108.
- 312 13. Chankong T, Theera-Umpon N, Auephanwiriyaikul S. Automatic cervical cell
313 segmentation and classification in Pap smears. Comput Methods Programs Biomed.
314 2014 Feb;113(2):539–56.
- 315 14. Zhang L, Kong H, Ting Chin C, Liu S, Fan X, Wang T, et al. Automation-assisted cervical
316 cancer screening in manual liquid-based cytology with hematoxylin and eosin
317 staining. Cytom Part A. 2014 Mar;85(3):214–30.
- 318 15. Mariarputham EJ, Stephen A. Nominated texture based cervical cancer classification.
319 Comput Math Methods Med [Internet]. 2015 [cited 2020 May 16];2015. Available
320 from: <https://www.hindawi.com/journals/cmmm/2015/586928/>
- 321 16. Zhao L, Yin J, Yuan L, Liu Q, Li K, Qiu M. An efficient abnormal cervical cell detection
322 system based on multi-instance extreme learning machine. In: Falco CM, Jiang X,
323 editors. Ninth International Conference on Digital Image Processing (ICDIP 2017)
324 [Internet]. SPIE; 2017 [cited 2021 Feb 12]. p. 104203U. Available from:
325 <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2281648>
- 326 17. Plissiti ME, Dimitrakopoulos P, Sfikas G, Nikou C, Krikoni O, Charchanti A. Sipakmed: A
327 New Dataset for Feature and Image Based Classification of Normal and Pathological
328 Cervical Cells in Pap Smear Images. In: Proceedings - International Conference on
329 Image Processing, ICIP. IEEE Computer Society; 2018. p. 3144–8.
- 330 18. Win KP, Kitjaidure Y, Hamamoto K, Myo Aung T. Computer-Assisted Screening for

- 331 Cervical Cancer Using Digital Image Processing of Pap Smear Images. Appl Sci
 332 [Internet]. 2020 Mar 5 [cited 2021 Feb 12];10(5):1800. Available from:
 333 <https://www.mdpi.com/2076-3417/10/5/1800>
- 334 19. Nirmal Jith OU, Harinarayanan KK, Gautam S, Bhavsar A, Sao AK. DeepCerv: Deep
 335 Neural Network for Segmentation Free Robust Cervical Cell Classification. In: Lecture
 336 Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence
 337 and Lecture Notes in Bioinformatics) [Internet]. Springer Verlag; 2018 [cited 2021 Feb
 338 12]. p. 86–94. Available from: [https://link.springer.com/chapter/10.1007/978-3-030-](https://link.springer.com/chapter/10.1007/978-3-030-00949-6_11)
 339 [00949-6_11](https://link.springer.com/chapter/10.1007/978-3-030-00949-6_11)
- 340 20. Gautam S, K. HK, Jith N, Sao AK, Bhavsar A, Natarajan A. Considerations for a PAP
 341 Smear Image Analysis System with CNN Features. arXiv [Internet]. 2018 Jun 23 [cited
 342 2021 Feb 12]; Available from: <http://arxiv.org/abs/1806.09025>
- 343 21. Zhang L, Lu L, Nogues I, Summers RM, Liu S, Yao J. DeepPap: Deep convolutional
 344 networks for cervical cell classification. IEEE J Biomed Heal Informatics. 2017 Nov
 345 1;21(6):1633–43.
- 346 22. Yilmaz A, Demircali AA, Kocaman S, Uvet H. Comparison of Deep Learning and
 347 Traditional Machine Learning Techniques for Classification of Pap Smear Images. 2020
 348 Sep 11 [cited 2021 Feb 12]; Available from: <http://arxiv.org/abs/2009.06366>
- 349 23. Ghoneim A, Muhammad G, Hossain MS. Cervical cancer classification using
 350 convolutional neural networks and extreme learning machines. Futur Gener Comput
 351 Syst. 2020 Jan 1;102:643–9.
- 352 24. Taha B, Dias J, Werghi N. Classification of cervical-cancer using pap-smear images: A

convolutional neural network approach. In: Communications in Computer and
 Information Science [Internet]. Springer Verlag; 2017 [cited 2021 Feb 12]. p. 261–72.
 Available from: https://link.springer.com/chapter/10.1007/978-3-319-60964-5_23

25. Keras Applications [Internet]. [cited 2021 Feb 12]. Available from:
<https://keras.io/api/applications/#available-models>

26. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale
 Visual Recognition Challenge. Int J Comput Vis [Internet]. 2014 Sep 1 [cited 2021 Feb
 12];115(3):211–52. Available from: <http://arxiv.org/abs/1409.0575>

27. Grandini M, Bagli E, Visani G. Metrics for multi-class classification: An overview
 [Internet]. arXiv. arXiv; 2020 [cited 2021 Feb 12]. Available from:
<http://arxiv.org/abs/2008.05756>

Figures

Figure 1

Common pipelines to classify CPS images: classification using hand-crafted features and classification using deep learning schemes.

Figure 2

Training accuracy (left) and training loss (right) of the architectures over 100 epochs.

Figure 3

Validation accuracy (left) and validation loss (right) of the architectures over 100 epochs.

Figure 4

Confusion matrix for classification of the test dataset using fine-tuned DenseNet169.

Figure 5

Sample images from the SIPaKMeD dataset: superficial-intermediate (a), parabasal (b), koilocytotic (c), metaplastic (d) and dyskeratotic (e) cells.

Figure 6

The general pipeline of the research project: image acquisition, pre-processing, feature extraction and classification.