

Canine Semen Evaluation using Transfer Learning Models

Jessica S. Velasco¹, Maria Victoria C. Padilla¹, Nilo M. Arago¹, Edev Paul E. De Vera¹, Flynn Ernest M. Domingo¹,
Roniel Emerson R. Ramos¹

¹Department of Electronics Engineering, Technological University of the Philippines, Manila, Philippines

ABSTRACT

This paper has experimented different architectures of transfer learning models (TLM) for canine semen evaluation and improving the performance of Deep Learning algorithms. There are 332 total datasets used for training and testing of different convolutional network models. These models are: MobileNetV1, InceptionV2, ResNet50, ResNet101 and Inception_ResNetV2. The canine sperm classification is successfully executed. In the evaluation of these models, a comparison of the manual method to the systems accuracy and error was made. Through these evaluations, the highest accuracy of 87% is obtained by Inception_ResNetV2 and the lowest accuracy of 40% is the MobileNetv1.

Key words : Convolutional Neural Network; Deep Learning; Canine Semen; Transfer Learning

1. INTRODUCTION

According to studies, there are three categories why canines is suffering infertility. Failure to copulate or ejaculate is the first one, next is the quality of semen is poor because of stress, environment factors etc. and lastly the prostatic diseases. Focusing on the poor semen quality, this is because of low sperm count, unusual motility and abnormal morphology. The possible cause of the deprived quality of canine semen are hormonal imbalance, medications, fatal fevers, and numerous diseases on the canine's testicles. In validation of the canine's fertility, it will undergo under numerous series of testing and examination to know the roots of the said problem. Complete blood analysis, urinalysis, canine semen testing, ultrasound, hormone testing and other testing are the sample of this laboratory examination [1].

There are many reasons for canine semen evaluation. Mainly, it is done to determine the fertility of canines prior to mating or breeding. The canine semen undergoes different test for the parameters namely the sperm concentration, motility and morphology. The data results are compared to the acknowledged normal parameters. The sperm's concentration must be above 70%, the sperm motility must be atleast 10 million sperm/ pound (body weight) and in the morphology must have greater than 70% normal sperms [2].

Sperm cells of canine are similar to cattle, human and bulls. Since it is similar to humans, we can implement this study in analyzing human semen.

2. MICROSCOPIC ANALYSIS

2.1 Microscopic Analysis

Microscopic analysis is performed as well as macroscopic analysis. Both are very important in Canine semen analysis. Therefore, condition identifies the degree or result of analysis.

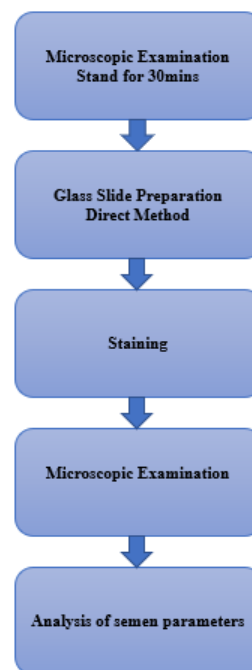


Figure 1: Steps in preparation for Semen Analysis

The preparation for analysis of semen is shown in Figure 1. In about 30 minutes, the semen must be in liquid form. Only 50µL is pipetted to a slide. The coverslip is placed inspected with staining for morphology and without staining for sperm concentration and motility.

2.2 Pathological Components

2.2.1 White Blood Cells (WBC) or Pus Cells

In a real-world examination, numerous of WBC in the semen, which could make the quality of sperm poor and damage the DNA or the genetic material [4].

2.2.2 Abnormal Sperm

More numerous abnormal sperms than normal sperm could show infertility of canines. Abnormal sperm have flaw on its head shape or tails, and it can be seen on high-power field (hpf) [5].

2.3 Microscopic Images

The datasets used in this study are images taken from the microscope. Figure 2 shows samples of microscopic images: Normal sperm, abnormal sperm and White Blood Cells (WBC).

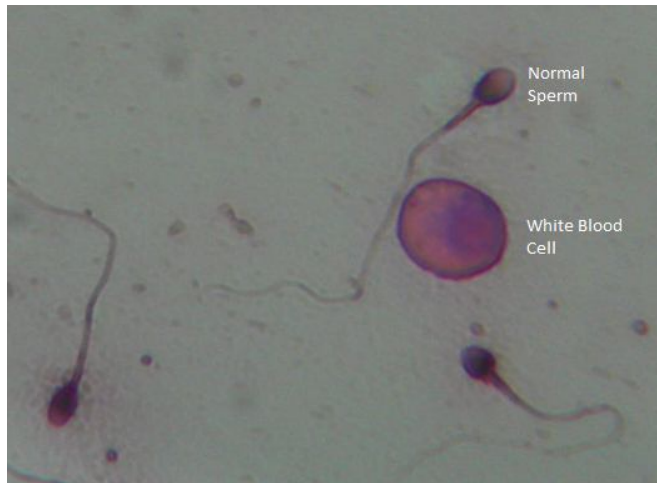


Figure 2: Sample image in Dataset

3. OBJECTIVES

3.1 General Objective

This study will develop an automated semen analyzer that will classify canine semen through Transfer Learning Models.

3.2 Specific Objectives

3.2.1 To semi-automate the microscope mechanism to be used for canine semen analysis.

3.2.2 To test and validate the best Transfer Learning Models (TLM) and compare the accuracy of the system to the conventional method of counting.

3.2.3 To utilize TLM in determining sperm cells and WBC counts through image and video processing.

4. LITERATURE REVIEW

The transfer learning method is a training the initial network. initial n layers are imitated to the residual n layers of a target network. The residual layers of target training are initialized randomly. It is now trained towards the target task. The errors are backpropagated to the initial features for fine-tuning. The

size of target database depends on the initial n layers of the fine-tuned target networks [6],[7],[8],[9].

The study used pre-trained ConvNet models as: MobileNetV1, InceptionV2, ResNet50, ResNet101 and Inception_ResNetV2.

4.1 MobileNetV1

It has mono filter to each input channel and pointwise convolution that is applied to a 1x1 convolution to combine the outputs of convolutional blocks, it also has a acknowledge convolution process which filters the input and combines into a new set of data outputs in a step. It consists of two layers which job is to filter and combine datasets [10].

4.2 Inception V2

It used 3x3 convolution, there are 3 traditional inception modules: 288 filters for each layer with 35x35 blocks, using reduction technique that has 768 filters and 17x17 blocks and a grid reduction technique with 8x8x1280 blocks. It is developed by Christian Szegedy and Zbigniew Wojna [11].

4.3 ResNet50

This convolutional network has a depth of 152 layers, which is 8 times better than VGG net. It is developed by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun [12].

4.4 ResNet101

The feature of extraction has 36 convolutional network. These are formed into 14 modules, these modules have linear residual connections forming it, excluding the first and last modules [12].

4.5 Inception-Resnet-V2

This is the structured of three residual layers and one Inception V4. It is developed by Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke and Alex Alemi [13].

5. METHODOLOGY

The system model of this study is shown in figure 3 below.

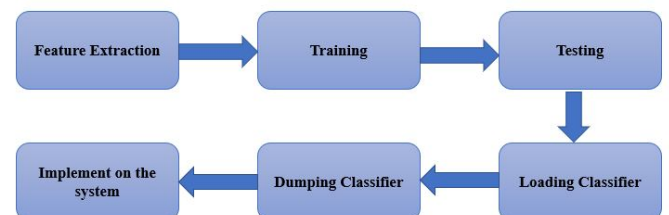


Figure 3: System Model

5.1 System Model

The system model of this study starts at loading the datasets to the model. The Model used in the study are: MobileNetV1, InceptionV2, ResNet50, ResNet101 and Inception_ResNetV2. The datasets have 3 categories, these are: Abnormal sperm, normal sperm and WBC. The network is trained and tested for classification of the datasets according to respective category. After the testing, the classifier is loaded to the system and used as classifier of category. Each category was evaluated per transfer learning model.

5.2 Data Collection

The microscopic images of the semen specimens are manually collected using a binocular microscope with 4x and 10x magnifier. Through the desired magnifier, it has high-power field (hpf) and low-power field (lpf). There are 332 total images and the breakdown is shown in Table 1.

Table 1: Dataset

Semen Sample	Number of Images
White Blood Cells	66
Abnormal Sperm	119
Normal Sperm	147
Total	332

The input images differ for pre-trained model. The common shape of the pre-trained model is a rectangular prism. This is equal to the product of length, height and depth. The input image is equal to the product of the of image and channel number. The number of channels is 3 for colored images and 1 for black and white images. Table 2 shows the size of input images for each model.

Table 2: Input for each model

Model	Input Image
MobileNetV1	720 x 1280 x 3
InceptionV2	720 x 1280 x 3
ResNet50	720 x 1280 x 3
ResNet101	720 x 1280 x 3
Inception-Resnet-V2	720 x 1280 x 3

5.3 Experiment: Feature Extraction, Training and Test

After gathering the 332 datasets, feature is extracted. The images were divided to 80% for training and 20% for testing. There are 299 training data, 33 testing data and 332 image datasets. The experiments were performed in an Intel Core i3 Processor workstation with GeForce GTX 650 GPU, 1GB at 1350 MHz. Tensorflow was utilized to train models [14][15][16][17]. Proponents limit the epoch for 4.1k. The figure below shows the sample training of pre-trained model.

```

Anaconda Prompt - python train.py --logtostderr --train_dir=training/ --pipelin...
Flow.python.training.supervisor is deprecated and will be removed in a future v
Instructions for updating:
Please switch to tf.train.MonitoredTrainingSession
2018-10-02 09:32:21.496849: I C:\tf_jenkins\workspace\rel-win\M\windows\PY35\te
nsorflow\core\platform\cpu_feature_guard.cc:137] Your CPU supports instructions
that this TensorFlow binary was not compiled to use: AVX
INFO:tensorflow:Restoring parameters from training/model.ckpt-14650
INFO:tensorflow:Starting Session.
INFO:tensorflow:Starting Session.
INFO:tensorflow:Saving checkpoint to path training/model.ckpt
INFO:tensorflow:Saving checkpoint to path training/model.ckpt
INFO:tensorflow:Starting Queues.
INFO:tensorflow:Starting Queues.
INFO:tensorflow:global_step/sec: 0
INFO:tensorflow:global_step/sec: 0
INFO:tensorflow:Recording summary at step 14650.
INFO:tensorflow:Recording summary at step 14650.
INFO:tensorflow:global step 14651: loss = 0.4326 (74.626 sec/step)
INFO:tensorflow:global step 14651: loss = 0.4326 (74.626 sec/step)
INFO:tensorflow:global step 14652: loss = 0.1988 (27.658 sec/step)
INFO:tensorflow:global step 14652: loss = 0.1988 (27.658 sec/step)
INFO:tensorflow:global step/sec: 0.0166314
INFO:tensorflow:global step/sec: 0.0166314
INFO:tensorflow:Recording summary at step 14652.
INFO:tensorflow:Recording summary at step 14652.
INFO:tensorflow:global step 14653: loss = 0.1288 (14.673 sec/step)
INFO:tensorflow:global step 14653: loss = 0.1288 (14.673 sec/step)
INFO:tensorflow:global step 14654: loss = 0.2413 (10.548 sec/step)
INFO:tensorflow:global step 14654: loss = 0.2413 (10.548 sec/step)
INFO:tensorflow:global step 14655: loss = 0.2919 (10.563 sec/step)
INFO:tensorflow:global step 14655: loss = 0.2919 (10.563 sec/step)
INFO:tensorflow:global step 14656: loss = 0.1645 (10.548 sec/step)
INFO:tensorflow:global step 14656: loss = 0.1645 (10.548 sec/step)
INFO:tensorflow:global step 14657: loss = 0.0918 (10.530 sec/step)
INFO:tensorflow:global step 14657: loss = 0.0918 (10.530 sec/step)
INFO:tensorflow:global step 14658: loss = 0.0374 (10.438 sec/step)
INFO:tensorflow:global step 14658: loss = 0.0374 (10.438 sec/step)
INFO:tensorflow:global step 14659: loss = 0.1255 (10.470 sec/step)
INFO:tensorflow:global step 14659: loss = 0.1255 (10.470 sec/step)
INFO:tensorflow:global step 14660: loss = 0.2390 (10.454 sec/step)

```

Figure 4: Sample training for pre-trained model

5.4 Data Gathered

Data were gathered with the help of a Registered Medical Technologist (RMT). Each data will be compared to the RMT's analysis and pre-trained model. The figure below is the data gathered for sperm morphology.

Data Analysis						
Sperm Morphology (abnormal)						
Sample	Automated					Manual
	MobileNetv1	Inceptionv2	ResNet50	ResNet101	Inception_Resnetv2	
1	30	12	4	8	8	9
2	4	10	4	6	7	5
3	50	1	9	1	1	4
4	10	11	3	5	6	3
5	21	8	6	7	6	5
6	20	5	1	8	10	8
7	35	2	2	7	7	4
8	9	1	5	7	6	8
9	40	2	7	2	2	3
10	10	10	6	6	4	5
11	21	11	1	5	5	9
12	30	9	3	6	6	3
13	29	2	8	3	3	9
14	1	6	3	3	2	4
15	20	10	0	5	4	7

Figure 5: Data gathered for sperm morphology - Abnormal

6. RESULTS

To achieve the objectives of the study, there were two main statistical tools used. First, the Friedman Rank Sum Test which was used to determine whether there are differences among treatments (Manual, MobileNetv1, Inceptionv2, ResNet50, ResNet101, Inception_ResnetV2) and the Nemenyi test.

6.1 Dependent Variable: Sperm Count

Table 3 and Figure 6 show that the median (bold line in the figure) sperm counts of Inceptionv2, ResNet50, ResNet101, and Inception_ResnetV2 are almost equal to the median of sperm counts in the Manual method which is 95.5. Meanwhile, the MobileNetv1 obtained a median of 30 which is relatively far from that of the manual.

Table 3: Summary statistics of the sperm count

Treatment	Min	Med	Max
Manual	61	95.5	126
MobileNetv1	18	30	56
Inceptionv2	54	99.5	138
ResNet50	54	90	131
ResNet101	58	99.5	131
Inception_ResnetV2	59	100	133

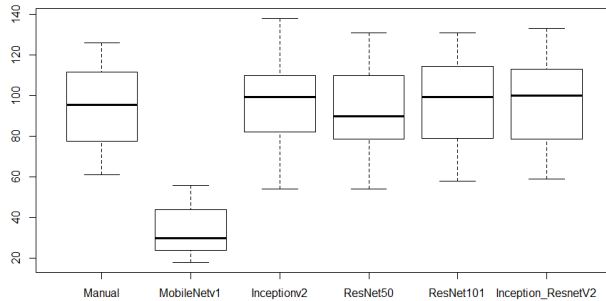


Figure 6: Box plot of the sperm count

The Friedman Test was performed, and it obtained a p-value of 2.2×10^{-16} , which is less than $\alpha=0.05$. Therefore, we have sufficient evidence to say that at least one of the sperm counts is significantly different among the treatments. Now in order to identify which treatment/s is/are significantly different, a post-hoc pairwise comparison was made using the Nemenyi Test. The following is the result as shown in Figure 7 and Table 4.

`Friedman rank sum test`

```
data: sperm_count, treatment and experimental_unit
Friedman chi-squared = 137.4, df = 5, p-value < 2.2e-16
```

Figure 7: Friedman test and Nemenyi test (RStudio) – Sperm Count

Table 4: Nemenyi test results for sperm count.

Treatment	p-value	Decision
Manual vs. MobileNetv1	7.1×10^{-14}	Significantly different
Manual vs. Inceptionv2	1.00	Not Significantly different
Manual vs. ResNet50	1.00	Not Significantly different
Manual vs. ResNet101	0.37	Not Significantly different
Manual vs. Inception_ResnetV2	0.93	Not Significantly different

At level of significance 0.05, while Manual and *MobileNetv1* obtained a significant difference, there is a sufficient evidence to say that there is no statistically significant difference in the between the Manual method and the *Inceptionv2*, *ResNet50*, *ResNet101*, and *Inception_ResnetV2*. That is, these four automated techniques yielded a promising result since its outcome is comparable to that of the Manual technique in identifying the sperm count.

6.2 Dependent Variable: Abnormal Sperm Morphology Count

Table 5 and Figure 7 show that the median (bold line in the figure) sperm counts of Inceptionv2, ResNet50, ResNet101, and Inception_ResnetV2 are almost equal to the median of sperm counts in the Manual method which is 5. Meanwhile, the MobileNetv1 obtained a median of 14 which is relatively far from that of the manual.

Table 5: Summary statistics of the sperm count

Treatment	Min	Med	Max
Manual	0	5	9
MobileNetv1	0	14	50
Inceptionv2	0	5	12
ResNet50	0	4	9
ResNet101	0	5	9
Inception_ResnetV2	0	4	10

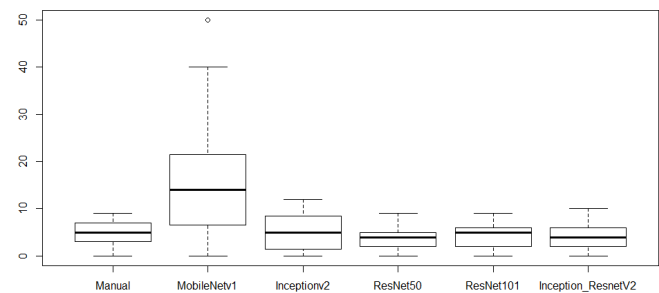


Figure 7: Box plot of the abnormal sperm morphology counts across treatments.

The Friedman Test was performed, and it obtained a p-value of 7.832×10^{-11} , which is less than $\alpha=0.05$. Therefore, we have sufficient evidence to say that at least one of the abnormal sperm counts is significantly different among the treatments. Now in order to identify which treatment/s is/are significantly different, a post-hoc pairwise comparison was made using the Nemenyi Test. The following is the result as shown in Figure 8 and Table 6.

`Friedman rank sum test`

```
data: abnormal_count, treatment and experimental_unit
Friedman chi-squared = 56.08, df = 5, p-value = 7.823e-11
```

Figure 10: Friedman test and Nemenyi test (RStudio) – Abnormal Morphology

Table 6: Nemenyi test results for abnormal sperm count morphology.

Treatment	p-value	Decision
Manual vs. MobileNetv1	1.8×10^{-5}	Significantly different
Manual vs. Inceptionv2	1.00	Not Significantly different
Manual vs. ResNet50	0.86	Not Significantly different
Manual vs. ResNet101	0.95	Not Significantly different
Manual vs. Inception_ResnetV2	0.84	Not Significantly different

At level of significance 0.05, while Manual and *MobileNetv1* obtained a significant difference, there is a sufficient evidence to say that there is no statistically significant difference in the between the Manual method and the *Inceptionv2*, *ResNet50*, *ResNet101*, and *Inception_ResnetV2*. That is, these four automated techniques yielded a promising result since its outcome is comparable to that of the Manual technique in identifying the abnormal sperm count morphology.

6.3 Dependent Variable: Presence of White Blood Cell

It could be observed on Table 7 that there are limited values for the counts, specifically, 0, 1, and 2. In testing whether the treatments are the same in correctly identifying the presence or absence of white blood cell in the samples, the McNemar's Testis more appropriate statistical tool to be used as shown in Table 8. This test measures the consistency in responses (presence or absence of WBC) across two treatments. The null hypothesis is that, the white blood cell is equal between Treatment A and Treatment B.

Table 7: Frequency of white blood cell counts across treatments.

Treatment	WBC Count		
	0	1	2
Manual	40	16	4
MobileNetv1	41	15	4
Inceptionv2	41	13	6
ResNet50	37	17	6
ResNet101	39	19	2
Inception_ResnetV2	39	17	4

Table 8: Frequency of white blood cell counts across treatments.

Treatment	p-value	Decision
Manual vs. MobileNetv1	0.8415	Not Significantly different
Manual vs. Inceptionv2	0.8348	Not Significantly different
Manual vs. ResNet50	0.5127	Not Significantly different
Manual vs. ResNet101	0.6547	Not Significantly different
Manual vs. Inception_ResnetV2	0.5637	Not Significantly different

Results lead to failure of rejecting the null hypothesis. Therefore, we have sufficient evidence to say that the determination of white blood cell is the same between the manual method and across treatment groups.

7. CONCLUSION

The application of deep convolutional neural network in microscopic images is applied. This also answer proponent's objective for classifying semen sample with sufficient accuracy using the transfer learning models. In this study, the microscopic images is taken manually and pre-labeled by RMT as: Abnormal Sperm, Normal Sperm and White Blood Cells (WBC). The best classifiers and the highest true positive

recall among the model is Inception_ResnetV2. The highest accuracy of 87% is obtained by Inception_ResNetV2 while the lowest accuracy of 40% is the MobileNetv1.

8. FUTURE WORK

More pre-trained models (TLM) and better specification of camera will be compared for a higher accuracy on classifying and counting of sperm.

9. ACKNOWLEDGEMENT

The authors would like to thank Technological University of the Philippines – Manila, RMTs (Jenelle M. Castillo, RMT and Vincent Joharra M. Abo, RMT) and Statistician (Ira Joy Martin) who helped us and for their untiring support.

REFERENCES

1. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. **Inception-v4, inception-resnet and the impact of residual connections on learning**, in *Thirty-first AAAI conference on artificial intelligence*, 2017.
2. D. Catherine Barnette, **Infertility in Male Dogs**.
3. C. Lopate, **Canine Semen Evaluation**, vol. I, p. 7, 2015.
4. N. Aziz, **The importance of semen analysis in the context of azoospermia**, *Clinics*, vol. 68, pp. 35-38, 2013.
[https://doi.org/10.6061/clinics/2013\(Sup01\)05](https://doi.org/10.6061/clinics/2013(Sup01)05)
5. E. E. Wallach and H. Wolff, **The biologic significance of white blood cells in semen**, *Fertility and sterility*, vol. 63, no. 6, pp. 1143-1157, 1995.
6. H. E. Chemes and V. Y. Rawe, **Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men**, *Human Reproduction Update*, vol. 9, no. 5, pp. 405-428, 2003.
<https://doi.org/10.1093/humupd/dmg034>
7. S. Inthiyaz, M. V. D. Prasad, R. Usha Sri Lakshmi, N. T. B. Sri Sai, P. P. Kumar, and S. H. Ahammad, **Agriculture Based Plant Leaf Health Assessment Tool: A Deep Learning Perspective**, *International Journal of Emerging Trends in Engineering Research*, vol. 7, no. 11, pp. 690-694, 2019.
<https://doi.org/10.30534/ijeter/2019/457112019>
8. A. D. M. Africa, P. B. T. Arevalo, A. S. Publico, and M. A. A. Tan, **A Fuzzy Neural Control System**, *International Journal of Emerging Trends in Engineering Research*, vol. 7, no. 9, pp. 323-327, 2019.
<https://doi.org/10.30534/ijeter/2019/15792019>
9. J. C. V. Puno, et al., "Quality Assessment of Mangoes using Convolutional Neural Network," in **2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)**, pp. 491-495, 2019.
10. H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. **Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset**

- Characteristics and Transfer Learning**, *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.
<https://doi.org/10.1109/TMI.2016.2528162>
11. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**, 2017.
 12. B. Raj, **A Simple Guide to the Versions of the Inception Network**, 2018.
 13. T. Akiba, S. Suzuki and K. Fukuda, **"Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes**, in *NIPS'17 Workshop: Deep Learning at Supercomputer Scale*, 2017.
 14. L. K. S. Tolentino, R. O. Serfa Juan, A. C. Thio-ac, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, **Static Sign Language Recognition Using Deep Learning**, *International Journal of Machine Learning and Computing*, vol. 9, no. 6, 2019.
<https://doi.org/10.18178/ijmlc.2019.9.6.879>
 15. J. Velasco, C. Pascion, J. W. Alberio, J. Apuang, J. S. Cruz, M. A. Gomez, B. J. Molina, L. Tuala, A. Thio-ac, and R. J. Jorda, **A Smartphone-Based Skin Disease Classification Using MobileNet CNN**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, 2019.
<https://doi.org/10.30534/ijatcse/2019/116852019>
 16. J. S. Velasco, M. K. Cabatuan, and E. P. Dadios, **Urine Sediment Classification Using Deep Learning**, in *Lecture Notes on Advanced Research in Electrical and Electronic Engineering Technology*, pp. 180-185, 2019.
 17. J. C. Puno, S. Rabano, J. S. Velasco, M. K. Cabatuan, E. Sybingco, and E. Dadios, **4-Band Resistor Recognition Using LeNet-5**, in *Lecture Notes on Advanced Research in Electrical and Electronic Engineering Technology*, pp. 97-102, 2019.