Machine Learning Aided Classification of Noise Distribution in Scanning Electron Microscopy Images

Sheikh Shah Mohammad Motiur Rahman, Michel Salomon, Sounkalo Dembélé
Université de Franche-Comté, CNRS
Institut FEMTO-ST
F-90000 Belfort and F-25000 Besançon, France
{sheikh.rahman,michel.salomon,sounkalo.dembele}@femto-st.fr

***Abstract*—Noise estimation is a crucial part of any modern supervised denoiser. Various statistical approaches are studied to estimate the noise, but generally these depend on a manual analysis of the images. To remove the manual dependency, it is important to automate the noise estimation process. In this paper, the initial phase of noise estimation, which is to identify the types of noise distribution, is performed using machine learning (ML) techniques. To make the images workable with ML techniques, a feature extraction process was performed. Hu's moment invariants, Haralick's texture, and color histogram are extracted from the images and stacked horizontally by scaling with MinMax scale the three features into one. Label encoder is used for normalizing the labels. Multiple ML techniques are trained and validated, and then tested with unknown images. The result is that stacking multiple ML techniques can produce better results with an accuracy above 90%. Stacking with the test set produces the following scores for Precision, Recall, and F1-score: 0.98, 0.88, 0.93, and 0.89, 0.98, 0.93 for Gaussian and Poisson respectively, with an average precision of 88%. These promising results prove the capability of ML techniques for image noise classification tasks where noise is artificially added. However, in real case, i.e. when the images come from a Zeiss Auriga FE SEM, which is the initial target, the classification is not as efficient. Thus, it is not always possible to work in a real denoising scenario if the model is trained with synthetic data.**

***Index Terms*—Noise estimation, Machine Learning, Ensemble Learning, Classification**

I. Introduction

The identification of noise in visual data is of great interest to the fields of image processing and computer vision. As a result, several distinct denoising algorithms have been created, and there is still a great deal of ongoing research on image denoising. However, the automated task of identifying noise types in images receives relatively little attention [1], [2]. A few methods are known with notable performance in the well-established research area of noise estimation and removal. Prior to image denoising, identifying the noise information in the image is a crucial step [3]. If the noisy image information is correctly recognized, the denoising process will be effective. There is not much academic work that addresses this issue, and many papers only address denoising problems, but the prior information to make deep learning based denoisers or blind denoisers still needs more attention.

In fact, a denoiser's goal is to create a denoised replica D_i of an original clean image O_i from an observation O_b that is thought to be noisy. A noise function N generates O_b with the formula $O_b = N(O_i)$. To represent N, there are a plethora of noise models available [4] including Gaussian, Poisson, Bernoulli, Speckle or Uniform noise and many more. Although denoisers are improving in terms of noise elimination level [6], the majority of published approaches are made for and tested against a certain main noise distribution (i.e. respecting a known distribution) [5]. Therefore, classification of noise information can be performed in an automated and accurate manner using machine learning methods, and it may also be possible to improve denoising if the classifier model is properly trained [3].

In [5], the authors mention several benefits of understanding the different kinds of noise in an image. Namely, recognizing the main noises and creating a standard denoiser library to solve any noise removal problem by decomposing the mixed denoising problem into its fundamental problems. In summary, the distribution of the detected noise type can help in creating a denoising pipeline. Moreover, noise type estimation is crucial in our case, as we are working with scanning electron microscopy images that have both Gaussian ([7]) and Poisson ([8]) noises. Thus, distinguishing the type of noise is an important task to perform before denoising these images.

Additionally, relevant research demonstrates that the performance of cutting-edge denoising algorithms can decrease sharply when the noise parameters are wrong. Therefore, it is essential for image processing/analysis algorithms to estimate noise characteristics accurately [9].

In this study, we present an automated approach based on machine learning to classify noise types, either Gaussian or Poisson (as our target images may exhibit both types). This study is a large-scale assessment of classical machine learning approaches and ensemble techniques (including bootstrapping, boosting and stacking).

The main contributions of this study are as follows:

Thanks to EIPHI-Bourgogne Franche-Comté Region for funding the MEB-3D project (EIPHI Graduate School contract ANR-17-EURE-0002).

- ML techniques are vastly investigated in image classification tasks.
- Multiple types of features are extracted and concatenated to make image classification possible with ML techniques.
- Stacked generalization, a multi-layer approach, is also studied to reduce the error bias in the model.
- 10-fold cross validation has been considered during training of each model to ensure there is no overfitting.
- Noisy SEM images generated by artificially adding noise.
- Prior experiments on noise type classification of SEM images were performed for the first time.
- Training from synthetic noisy images may not yield the same results as with real noisy SEM images.
- An initiative for automatic noise type classification of SEM images.

The rest of the paper is organized as follows. Section II describes the background and related works. The process we performed during the analysis of the images is presented in Section III. Section IV describes the comparative results and findings in detail. Finally, Section V concludes the paper with possible future work.

II. Background and Related Works

Various aspects of image capture, such as acquisition, quantization, formatting, and compression, causes noise in the final image. For example, the Scanning Electron Microscopy (SEM) image capture process is divided into five stages and each stage is assumed to be Poisson distributed [8]. However, we found that it really depended on the setup and, more importantly, the scan speed. In order to perform denoising on SEM images, one must identify the types of noise and the level of noise. The process of gathering image noise information is known as noise estimation.

While noise estimates are rather uncommon, there is a vast and broad body of work on image denoising. To estimate noise, either a single image or a collection of photos can be employed. Overestimation issues exist in noise estimation from multiple photos [10]. The mean absolute deviation (MAD) is the foundation of a popular estimation technique [11]. The slope of the smooth or low-textured region is utilized to calculate the signal-dependent noise level for each intensity interval, according to [12]. Three methods were put forth by the authors of [13] to estimate noise levels using training samples and (Laplacian) statistics from actual image data.

Although overestimation problems in estimating noises from multiple images have been noticed by [10], the articles by [5], [14], [15] showed that it is possible to increase the denoising performance by estimating noise types using machine learning or deep learning classification techniques. Thus, we were inspired to continue the experiment by focusing specifically on noise type identification where these techniques focused on denoising.

Machine learning approaches [29] like K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART-Decision Tree), Naive Bayes (NB), Support Vector Machine (SVM), Gradient Boosting (GB), Hist Gradient Boosting (HisGB), Ada boosting (ADA), eXtreme Gradient Boost (XGB), Extremely Randomized Tree (ET), Random Forest (RF) and stacking [30] concept have enough fame in classification tasks. Usually, those techniques are widely used in tabular data classifications. However, it is possible to extract relevant features from images and reshape the features to apply above mentioned techniques to classify the noise types from images.

Hu Moments [21], Haralick Texture [18]–[20] and Color Histogram [22] are high level features of an image [16]. The defect extraction algorithm frequently uses the moment feature as a feature descriptor. Using the concept and properties of invariant moments, we address the issue that the seven feature components of Hu moments have large magnitude differences and are scale-dependent [17]. In order to measure the spatial relationship between adjacent pixels in an image, Haralick *et al.* [21] suggested using a gray-level co-occurrence matrix (GLCM). Due to their clarity and straightforward interpretations, Haralick texture characteristics, calculated from the GLCM, are frequently used and have been effectively applied in various classification tasks. The color histogram technique is the one that is most frequently used to derive an image's color characteristics. It depicts how an image's color bins are distributed in terms of frequency. Obviously, as we deal with gray level images, the color histogram is in fact a grayscale one (same value for the three color channels).

III. Methodology

The proposed and experimented methodology is depicted in Figure 1. However, before discussing the methodology, it is appropriate to explain the information about the dataset. Although Gaussian and Poisson noise can be exist in SEM images, we prepare a noisy dataset of scanning electron microscopy images. Both types of noise artificially added to the images from [23], denoted by $D1$, and few other noisy images collected from [24], denoted by $D2$. The training, validation, and test sets were then obtained by randomly selecting images in $D1$ and $D2$ such that 3,245 Gaussian and 3,218 Poisson images were used for training, 456 Gaussian and 457 Poisson images for validation, and 440 Gaussian and 442 Poisson images for testing. However, $D2$ has noisy images itself but the SEM images in $D1$ were artificially made noisy with different noise measurements. For Gaussian noise, the standard deviation (σ) values 20, 25, 30, 35, 40, 45 and 50 [25] were applied respectively. On the other hand for Poisson noise, the λ values 10, 20, 30, 40, 50, 35 and 45 [26] are used respectively.

The methodology starts with loading the dataset into the model with the pre-structure format which is done during the preparation and generation of the dataset. Then, image feature extraction and data labeling are performed.

A. Features Extractions and Data Labeling

The images in both datasets $D1$ and $D2$ had multiple dimensions. Thus, a resizing of all the images (600X400)

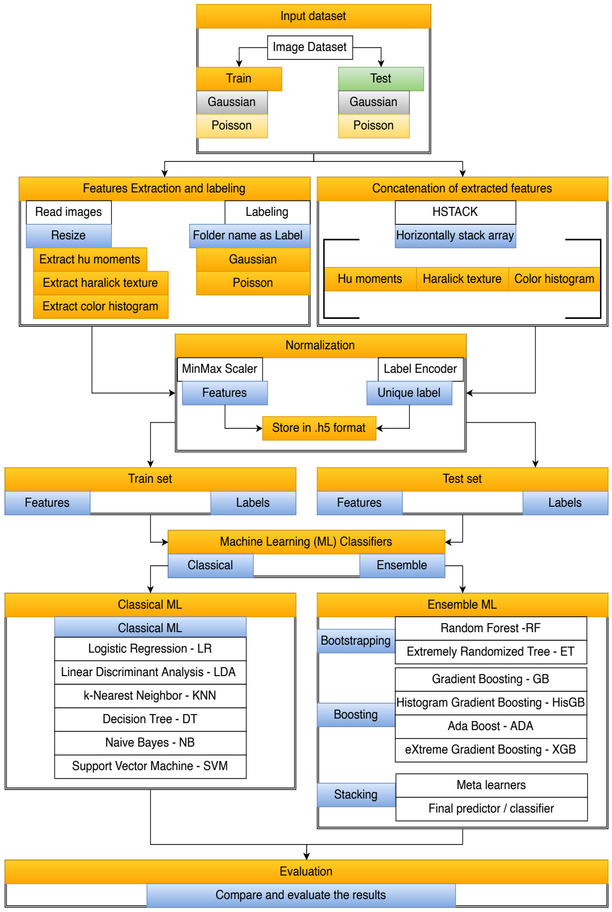

Figure 1. Proposed methodology.

is performed, then the Hu invariant moments, Haralick texture and color histogram are extracted from the images and concatenated horizontally (column by column). In parallel, the sub-folders (Gaussian or Poisson) from which the images come from are used as labels.

Few more information about the extracted features are described as follows:

- Hu moments invariant: Due to their invariance properties with respect to image translation, scaling, and rotation, moment invariants have been widely used for image pattern recognition in a number of applications [33].
- Haralick texture: Common texture descriptors used in picture analysis are Haralick texture characteristics. The gray levels of an image are reduced, a procedure known as quantization, in order to calculate the Haralick features.
- Color histogram: A visual depiction of how colors are distributed in an image. A histogram's data is derived by counting the instances of each color that could exist in the image according to the chosen color model [34].

B. Normalization

We have now stacked features in an array that is normalized by applying MinMax scaler and in the same time the unique labels (Gaussian and Poisson in our case) are encoded using label encoder. Finally, the normalized features and labeled are stored in .h5 format to further use for training and test (because the same process was followed to prepare the training and test data sets).

C. Machine Learning (ML) Classifiers

Classical machine learning classifiers are widely studied, as well as ensemble techniques such as bootstrapping, boosting and stacking, in order to identify the most efficient ones in image classification. Thus, Logistic Regression (LR), k-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART-Decision Tree), Naive Bayes (NB), Support Vector Machine (SVM), Gradient Boosting (GB), Histogram-based Gradient Boosting (HisGB), Ada boosting (ADA), eXtreme Gradient Boost (XGB), Extremely Randomized Tree (ET), Random Forest (RF) and Stacking of multiple meta learners are experimented in this study. We used k-fold cross validation [35] to overcome the overfitting problems in learning of classifiers.

Finally, the extracted and normalized features were used to train the ML techniques and perform an evaluation with the test set.

D. Performance Evaluation

The proposed methodology has been evaluated and validated with Confusion matrix, Precision, Recall, F1-score, Accuracy, ROC AUC, and Boxplots [30]. The most popular and straightforward method for assessing performance with two or more types of classes in any classification task is the confusion matrix. The confusion matrix is just a table with two dimensions—called real and predicted—that aid in making any forecast. In the confusion matrix there are four terms: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

All these performance indicators can be calculated using the following formulas:

- **Precision:** TP/(TP+FP).
- **Recall:** TP/(TP+FN).
- **F1-score:** $2\times$(Precision$\times$Recall)/(Precision+Recall).
- **Accuracy:** (TP+TN)/(TP+TN+FP+FN).
- **ROC AUC:** The Area Under the Curve (AUC), which serves as a summary of the ROC curve [32], is a measurement of a classifier's capacity to discriminate between classes. The performance of the model in differentiating between the Gaussian and Poisson classes improves with increasing AUC.
- **Boxplots** [31] are helpful for figuring out how data are distributed, spotting outliers, and contrasting distributions. It is simpler to see differences between distributions when the boxplot representation is slightly altered. They summarize the data concisely, and the placement of the box and whisker markers makes it simple to compare the classification accuracy.

More information on the performance indicators can be found in [27], [30].

IV. Results and Discussion

The experimental results of our proposal are presented comparatively in Table I, showing the ability of different ML and ensemble techniques to classify Gaussian and Poisson noises. From the values in the table, it can be seen that ensemble techniques perform better than classical ML techniques. Mostly, boosting techniques have significant results in terms of ROC AUC, Precision, Recall, F1-score and Accuracy. ADA, HisGB, GB, XGB and Stacking provide 93% of accuracy where RF and ET (ensemble techniques) provide 92% accuracy.

Table I
Comparison of the results of different ML and ensemble techniques

ML	*ROC AUC*	*Precision*	*Recall*	*F1-score*	*Accuracy*
LR	0.98	0.93	0.89	0.91	0.91
LDA	0.98	0.95	0.87	0.91	0.91
KNN	0.97	0.88	0.91	0.90	0.90
CART-DT	0.90	0.90	0.90	0.90	0.90
NB	0.81	0.73	0.99	0.84	0.81
SVM	0.97	0.94	0.85	0.90	0.90
RF	0.98	0.91	0.93	0.92	0.92
ET	0.98	0.91	0.92	0.92	0.92
GB	0.99	0.93	0.94	0.93	0.93
ADA	0.98	0.93	0.93	0.93	0.93
HisGB	0.99	0.92	0.94	0.93	0.93
XGB	0.98	0.92	0.93	0.93	0.93
Stacking	0.99	0.93	0.94	0.93	0.93

Stacked generalization (stacking) is known as a super learner and has the particularity of minimizing error bias. We can in particular observe that it produces 94% for recall (or sensitivity), 93% for precision and F1-score. Because of its overall performance, we chose stacking as the winning technique among all others to solve noise type classification problems from SEM images.

Several observations can be made from the the boxplots shown in Figure 2. First, it can be seen that RF has a symmetric distribution of accuracy achieved from 10-fold cross validation, with two outliers. Second, LR and LDA provide very stable results exhibiting a symmetric distribution of accuracy, with an average rate of 91%. Third, stacking and GB provide the best performance for the validation set in terms of all matrices, including the AUC ROC. Thus, the concept of stacking or GB may be useful for other experiments in the test set. However, GB is already employed within the Stacking as one of the meta-learners thus we count Stacking as winner in noise classification apart from other classifiers.

The number of correctly classified and misclassified test data samples using Stacking are shown in confusion matrix displayed in Figure 3. Hence, Stacking was able to distinguish Gaussian and Poisson images with 88% and 98% of accuracy respectfully. Overall, the proposed approach with Stacking provides 0.98, 0.88, 0.93 (Precision, Recall, F1-score) and 0.89, 0.98, 0.93 (Precision, Recall, F1-score) scores for Gaussian and Poisson noises respectively, where the average accuracy is around 93% (+/- 1).

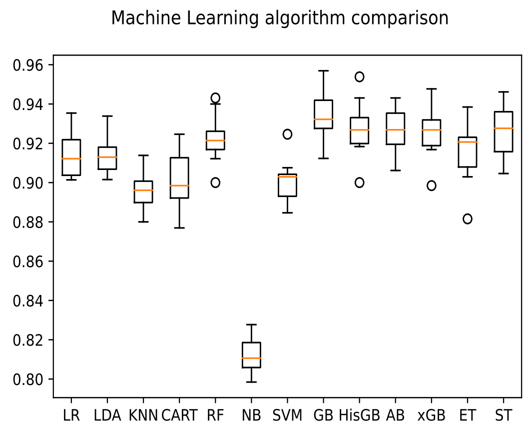

Figure 2. Comparison of ML and ensemble techniques accuracy.

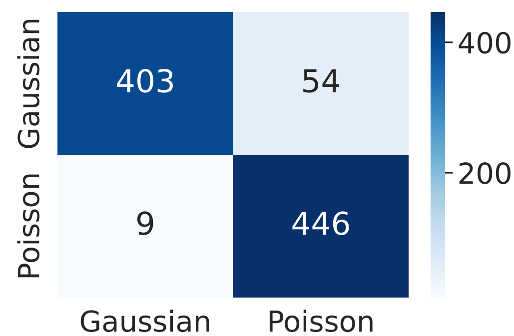

Figure 3. Confusion Matrix.

RF, CART-DT and GB are used as meta learners in the stack while XGB is used as the final classifier. A few other combinations have been tried, but this combination of classifiers provided the best performance.

The provided results show that artificially added noise or synthetic noisy images can be distinguishable using the proposed model. However, we tested it with the real noisy images collected from the Zeiss Auriga FE SEM of our lab (FEMTO-ST). Unfortunately, with real noisy images, the performance of the model with Gaussian noise was poor, since the model classified the images mainly as having Poisson noise. Therefore, it is crucial to create a real noisy dataset with manual noise labeling to improve model training. In addition, learning noise from more smoother pixels may improve the classification accuracy of real noisy images. This is because smoother pixels can provide more meaningful information about noise types than the whole images.

V. Conclusion and Future Work

Machine learning techniques are widely used and have shown better results in the field of classification, mainly on tabular data. In this study, these techniques are used and evaluated in image classification tasks. Specifically, the goal is to classify the types of noise existing in the images by targeting the automation of noise estimation using ML techniques. We found that the ensemble ML techniques, especially boosting and stacking techniques, outperformed in image classification. Noisy images (mixed SEM and optical) obtained by artificially adding noise were used to train and test the proposed approach. As a result, boosting and stacking provided the highest performance with 93% (+/- 1) of accuracy. However, our proposal is not properly able to distinguish the real noisy SEM images. Thus, a message to future researchers is not to use ML techniques for classification of real noisy images, as it may not work, especially for SEM images. However, deep learning (DL) techniques can be further investigated to see if DL can successfully distinguish the type of noise in images. Subsequently, other image features can be evaluated to make a more concrete decision about using ML in noise type classification.

ACKNOWLEDGMENT

This work was funded by 2020-21 EIPHI-Bourgogne Franche-Comté Region in the context of the MEB-3D project and supported by the EIPHI Graduate School (contract ANR-17-EURE-0002).

REFERENCES

[1] Kumar, Rakesh and Saini, B., 2012. Improved Image Denoising Technique Using Neighboring Wavelet Coefficients of Optimal Wavelet with Adaptive Thresholding. International Journal of Computer Theory and Engineering, pp.395-400.

[2] Tripathi, M., 2021. Facial image noise classification and denoising using neural network. Sustainable Engineering and Innovation, 3(2), pp.102-111.

[3] Kunaraj, K., Maria Wenisch, S., Balaji, S. and Mahimai Don Bosco, F.P., 2019, September. Impulse noise classification using machine learning classifier and robust statistical features. In International Conference On Computational Vision and Bio Inspired Computing (pp. 631-644). Springer, Cham.

[4] Boyat, A.K. and Joshi, B.K., 2015. A review paper: noise models in digital image processing. arXiv preprint arXiv:1505.03489.

[5] Lemarchand, F., Findeli, T., Nogues, E. and Pelcat, M., 2020, September. Noisebreaker: Gradual image denoising guided by noise analysis. In 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.

[6] Liu, P., Zhang, H., Zhang, K., Lin, L. and Zuo, W., 2018. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp.773-782).

[7] Marturi, N., Dembélé, S. and Piat, N., 2014. Scanning electron microscope image signal-to-noise ratio monitoring for micro-nanomanipulation. Scanning: The Journal of Scanning Microscopies, 36(4), pp.419-429.

[8] Timischl, F., Date, M. and Nemoto, S., 2012. A statistical model of signal–noise in scanning electron microscopy. Scanning, 34(3), pp.137-144.

[9] Tang, C., Yang, X. and Zhai, G., 2014. Noise estimation of natural images via statistical analysis and noise injection. IEEE Transactions on Circuits and Systems for Video Technology, 25(8), pp.1283-1294.

[10] Healey, G.E. and Kondepudy, R., 1994. Radiometric CCD camera calibration and noise estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(3), pp.267-276.

[11] Donoho, D.L., 1995. De-noising by soft-thresholding: IEEE Transactions on Information Theory.

[12] Förstner, W., 2000. Image preprocessing for feature extraction in digital intensity, color and range images. In Geomatic method for the analysis of data in the earth sciences (pp. 165-189). Springer, Berlin, Heidelberg.

[13] De Stefano, A., White, P.R. and Collis, W.B., 2004. Training methods for image noise level estimation on wavelet components. EURASIP Journal on Advances in Signal Processing, 2004(16), pp.1-8.

[14] Liu, F., Song, Q. and Jin, G., 2020. The classification and denoising of image noise based on deep neural networks. Applied Intelligence, 50(7), pp.2194-2207.

[15] Sil, D., Dutta, A. and Chandra, A., 2019, October. Convolutional neural networks for noise classification and denoising of images. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 447-451). IEEE.

[16] Nixon, M. and Aguado, A., 2019. Feature extraction and image processing for computer vision. Academic press.

[17] Ji, X., Guo, H. and Hu, M., 2019, September. Features extraction and classification of wood defect based on HU invariant moment and wavelet moment and BP neural network. In Proceedings of the 12th International Symposium on Visual Information Communication and Interaction (pp. 1-5).

[18] Porebski, A., Vandenbroucke, N. and Macaire, L., 2008, November. Haralick feature extraction from LBP images for color texture classification. In 2008 First Workshops on Image Processing Theory, Tools and Applications (pp. 1-8). IEEE.

[19] Vamsha Deepa, N., Krishna, N. and Hemanth Kumar, G., 2017, October. Feature extraction and classification of X-ray lung images using Haralick texture features. In International Conference on Next Generation Computing Technologies (pp. 899-907). Springer, Singapore.

[20] Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T. and Garpebring, A., 2019. Gray-level invariant Haralick texture features. PloS one, 14(2), p.e0212110.

[21] Haralick, R.M., Shanmugam, K. and Dinstein, I.H., 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), pp.610-621.

[22] Zhu, L.Q. and Zhang, Z., 2010, August. Auto-classification of insect images based on color histogram and GLCM. In 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (Vol. 6, pp. 2589-2593). IEEE.

[23] Aversa, R., Modarres, M.H., Cozzini, S., Ciancio, R. and Chiusole, A., 2018. The first annotated set of scanning electron microscopy images for nanoscience. Scientific data, 5(1), pp.1-10.

[24] Dibakar Sil, Arindam Dutta, Aniruddha Chandra, March 1, 2019, "CNN based noise classification and denoising of images", IEEE Dataport, doi: https://dx.doi.org/10.21227/3m26-dw82.

[25] Couturier, R., Perrot, G. and Salomon, M., 2018, December. Image denoising using a deep encoder-decoder network with skip connections. In International conference on neural information processing (pp. 554-565). Springer, Cham.

[26] Khademi, W., Rao, S., Minnerath, C., Hagen, G. and Ventura, J., 2021. Self-supervised poisson-gaussian denoising. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2131-2139).

[27] Rahman, S.S.M.M., Islam, T. and Jabiullah, M.I., 2020. PhishStack: evaluation of stacked generalization in phishing URLs detection. Procedia Computer Science, 167, pp.2410-2418.

[28] Wolpert, D.H., 1992. Stacked generalization. Neural networks, 5(2), pp.241-259.

[29] Muhammad, I. and Yan, Z., 2015. SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY. ICTACT Journal on Soft Computing, 5(3).

[30] Motiur Rahman, S.S.M. and Saha, S.K., 2018, December. StackDroid: Evaluation of a multi-level approach for detecting the malware on android using stacked generalization. In International Conference on Recent Trends in Image Processing and Pattern Recognition (pp. 611-623). Springer, Singapore

[31] McGill, R., Tukey, J.W. and Larsen, W.A., 1978. Variations of box plots. The american statistician, 32(1), pp.12-16.

[32] Narkhede, S., 2018. Understanding auc-roc curve. Towards Data Science, 26(1), pp.220-227.

[33] Huang, Z. and Leng, J., 2010, April. Analysis of Hu's moment invariants on image scaling and rotation. In 2010 2nd international conference on computer engineering and technology (Vol. 7, pp. V7-476). IEEE.

[34] Exarchos, T.P., Papadopoulos, A. and Fotiadis, D.I. eds., 2009. Handbook of research on advanced techniques in diagnostic imaging and biomedical applications. IGI Global.

[35] Refaeilzadeh, P., Tang, L. and Liu, H., 2009. Cross-validation. Encyclopedia of database systems, 5, pp.532-538.

[36] Amin Tavakoli and Ali Pourmohammad, 2012. Image Denoising Based on Compressed Sensing. International Journal of Computer Theory and Engineering, 4(2), pp.266-269.