Proc. of the 7th International Conference on Engineering and Emerging Technologies (ICEET)
27-28 October 2021, Istanbul, Turkey

Device Identification Using Discrete Wavelet Transform

Supriya Yadav
School of Engineering and Digital Arts
University of Kent
Canterbury, UK
sy227@kent.ac.uk

Pooja R.Khanna
School of Engineering and Digital Arts
University of Kent
Canterbury, UK
pk327@kent.ac.uk

Gareth Howells
School of Computing
University of Kent
Canterbury, UK
W.G.J.Howells@kent.ac.uk

***Abstract*—This paper investigates the effectiveness of employing measured hardware features mapped into the frequency domain for devices identification. The technique is to utilize Discrete Wavelet Transform (DWT) coefficients as distinguishing features. The DWT coefficients address the degree of relationship between the investigated features and the wavelet function at different occurrences of time. Therefore, DWT coefficients carry useful temporal information about the transient activity of the investigated wavelet features. We study the impacts of utilizing different wavelet functions (Coiflets, Haar and Symlets) on the performance of the device identification system. This system yields 92.5 % of accuracy using Sym6 wavelet. A comparison is made of the accuracy of wavelet features and raw features with standard classifiers.**

Keywords— Security, Device Authentication, Discrete Wavelet Transform, Multidimensional space.

I. INTRODUCTION

Nowadays, with constantly rising cyber scams, securing credentials has become a main focus for organizations throughout the world. In the domain of PKI, the real credential is the private key. The traditional way to protect the private key is through Hardware Security Modules (HSMs), Smartcards, TPMchips and Key Management. Of course, the security of HSM is very high (often FIPSlevel [1]). However, applications connect to the HSM via username and password and/or client certificates stored in .pfx/. p12 files [2] [3]. Often these can be hacked with simple social engineering tricks. Additionally, HSMs are relatively expensive [4]. In a post-pandemic workforce, remote connections have greatly increased, and this has created additional security concerns for CISOs that cannot be met with ever-tightening budgets. Large organizations have to choose between key protection and productivity. Unfortunately, today it is all too likely that organizations are relying on the underlying native security offered by a device's Operating System, the device hardware itself and the device microprocessor providers. The question is, do these go far enough?

This paper introduces a novel wavelet feature based multivariate Gaussian distribution classifier framework to identify device uniquely. We assessed the performance of the following classification algorithms: Logistic Regression, SVM and Multivariate Gaussian Distribution, where we evaluate and compare the performance of the proposed wavelet features and without wavelet feature based Multivariate Gaussian distribution with the standard classifiers.

The paper is organized as follows. Section II (A) introduces the novel technique and section II (B) presents an overview of some common approaches used and their limitations. Section III introduces an overview of the proposed system followed by experimental methodology and results in Section IV & Section V and Section VI concludes the paper.

II. BACKGROUND

This section introduces the concept of ICMetrics technology, which is used as the basis in our work to identify devices from its own operating characteristics [5].

Integrated Circuit Metrics (ICMetrics) is a software client which reads various dynamic and static hardware and software parameters in a device. The device characterizations employed by the system are known generically as features. Features are a major part of the ICMetrics system, and the features utilized straightforwardly influence the strength of the security provided [6][7]. Every time a sensitive operation (for example authentication) is required, ICMetrics reads these feature values, ensures it's the genuine device, reconstructs the digital fingerprint and completes the authentication successfully. On a rogue device, the digital fingerprint will obviously not match ICM expected value, hence the operation fails, thereby denying the opportunity to a potential bad actor. This process eliminates the possibility of both online and offline Brute Force attack [5].

An ICMetrics system generally consists of two phases, the *calibration phase* and the *operation phase*.

Calibration is useful for extracting suitable features in pre-production. Calibration is carried out once per application domain. Recording the features associated with a device depends on the nature of the device and what can be derived from it. The operation phase starts each time digital fingerprint is required for device authentication [8].

This describes the typical ICMetrics process for authenticating the devices. In this paper, we concentrate on the calibration phase of the system.

A. Previous Work/ Comparing Techniques

This section reviews existing device authenticating techniques.

978-1-6654-3897-1/21/$31.00 ©2021 IEEE

1) PKCS#12 Files

PKCS #12 defines an archive file format for storing many cryptography objects as a single file. It is commonly used to bundle a private key with its X.509 certificate. Assuming login credentials are compromised with a standard social engineering attack, malware can be deployed that then skims the private key associated with client certificate, so the hackers can access the network (e.g. HSM) or sensitive target keys. More than 2 decades ago, W32M Caligula malware was designed to infect a victim's machine, scan for PGP key rings and upload them silently to the hackers [9].

2) Multi Party Computation (MPC)

One of the leading MPC vendors, splits a secret in multiple shares, 1 is stored on an end-entity device and others at server level (on-prem/SaaS etc.). Every time a crypto operation is required, all the shares are combined and the operation is performed. To secure the end-entity share at device level, there is total dependence on native security features, all of which have been breached. For example, the Pegasus attack against WhatsApp encryption keys, an invisible zero-click exploit in iMessage or the Jeff Bezos iPhone hack. If the end-entity device is compromised by malware, it can potentially skim the new refreshed key/share even before the victim can use it. There is no way for the sever to know whether malware has taken over the device.

3) TPM Chips

Being hardware, if there is a vulnerability detected even in library implementation, the devices needs to be physically recalled –a logistical nightmare. For example, In October 2017, it was reported that a code library developed by Infineon, (used in its TPMs) contained a vulnerability, known as ROCA, which allowed RSA private keys to be inferred from public keys. As a result, all systems depending upon the privacy of such keys were vulnerable to compromise, such as identity theft or spoofing. Estonia paid the price and had to recall 750,000 ID cards. For large scale IOT devices (100K-200K or above), TPMs are not economically viable since they cost USD 4-5 per device.

4) SRAM PUF

A software client is deployed in each device. At silicon level, this client reads the unique submicron physical characteristics of a chip's SRAM and using this, generates an asymmetric key pair. A digital certificate is then issued for proof of identity. SRAM-based PUFs do not have anti-malware capabilities and are highly vulnerable to attacks by malware running on system's (micro) processor. It is the device owner's responsibility to protect the boot sequence. If a hardware vulnerability is ever discovered (as with Spectre, Meltdown, SGAxe and Crosstalk vulnerabilities), many keys will have to be revoked and renewed. Most importantly, SRAM PUF does not work on all devices; such as Intel, AMD processors and Apple –that's a significant chunk of enterprise security devices.

III. System Overview

This section introduces the characteristics of computing devices and classification based on device usage and its hardware. Our chosen platform is general purpose computing devices, as they have a wide range of applications which can be prone to attacks. Hence, we explore the distinctive device properties which can be used for device identification. We investigate a selection of device characteristics for their suitability to convey maximum information [7]. This selection criterion is introduced in Section B.

The working of this system is categorized in four stages

A. Criteria for Good Features- this portion of the paper introduces criteria for good features.
B. Feature capture – At this stage we collect feature data from the devices.
C. Feature selection- Out of all the data collected we select features meeting the required criteria.
D. Wavelet –map the features into the frequency domain generating wavelet coefficients subsequently employing these coefficients as features for classification

A. Criteria for Good Features

The requirements for good features are as follows:

1) Features that show high inter-sample variance means variation in measured values obtained from differing devices
2) Features that show low intra-sample variation means variation in measured values obtained from within a single device.
3) Correlated features provide a greater level of obfuscation.

We examined the possible overlap of the data between two or more devices based upon high inter-sample variance and low intra-sample variance [10].

In this experiment we used features that show low intra-sample variance and high inter-sample variance.

B. Feature Capture

Features are a main part of the ICMetric system. For uniquely device identification features are required to provide distinguishability for similar devices using the same features. The features not only have to provide sufficient variance but also the features should remain unknown to any unauthorized access, therefore the features need to come from a variety of sources on the device to prevent easy discovery of the features that are included.

In order to allow for a wide ranging set of features, the particular focus of this work has been on the more accessible iOS platform, due to the great variety of devices it provides, which allows for an in-depth analysis of how the features affect the system. Thus, the data was gathered from multiple devices in order to fully ascertain the range of each particular feature's values. The devices tested includes two different models of MacBook (4 per model, in total 8 devices) with different chip set. Additionally, several devices with identical chipsets of the same model were tested to obtain data from devices. The features that were looked at in-depth were narrowed down via their observed variations in value from a large selection of candidate features [10].

C. Feature Selection

In the previous section, we mentioned the criteria for good feature selection. In the previous work, we analyzed the features shown in Table I. These features exhibit multiple multimodal distribution from the collected data [11]. We concluded there were many challenges to model multimodal distributions. The challenge is to characterize the distribution properly without misrepresenting the data. Hence, our focus

in this work is to explore ways to model multimodal features effectively. Therefore, we investigate these features in the frequency domain to compare the performance based on classifiers accuracy.

In this study, features are extracted from each device. To select the most informative features, two different types of analysis were completed: inter-sample and intra-sample [5]. The main motivation is to identify those features that show high inter-sample and low intra-sample variation [12] [13]. After identification of the features, features fitting these criteria were selected and divided into the feature sets shown in the Table I below. This is also done to check the performance against different category of features.

There are three categories of features that are collected i.e., CPU related features, speed of hard disk related features and memory-based features. As they give a proper system profile and are easier to collect.

TABLE I. Shows list of potential features

Feature Set1	Maximum speed for copy function Maximum speed for scale function Maximum speed for add function Maximum speed for triad function Average duration for copy function Average duration for scale function Average duration for add function Average duration for triad function
Feature Set 2	Sequential write(block)%CPU Sequential write(block)MB/sec Sequential write(rewrite)%CPU Sequential write(rewrite) MB/sec Sequential read (per char) %CPU Sequential read (per char)MB/sec
Feature Set 3	Duration for add function Quickest duration for add function Longest duration for add function

D. Wavelets

Wavelets are functions that fulfill certain mathematical necessities and remain used in demonstrating data or additional functions. Wavelets are a family of simple functions that can be utilized to approximate other functions by extension in orthonormal arrangement [14]. One of the critical benefits of wavelets is their capacity to spatially accommodate features of a function like discontinuities and fluctuating frequency behavior. A wavelet transform is a lossless straight transformation of a signal or information into coefficients on a premise of wavelet function [15]. Performing the discrete wavelet transform (DWT) of a signal x is done by passing it through low pass filters (scaling functions) and high pass filters simultaneously [16].

The results provide the detail coefficients (from the high-pass filter) and approximation coefficients (from the low pass). The output of the low-pass filter is then subsampled by 2 and further processed by passing it again through a new low- pass filter and a high- pass filter with half the cut-off frequency of the previous one, This decomposition has halved the time resolution since only half of each filter output characterises the signal. Though, every output has half the frequency band of the input, so the frequency resolution has been doubled.

A two level DWT for N data. The Number of data is halved after every filtering and down sampling operation, this speed up the classification process. A wavelet transform is applied on the output of low pass filter [h(n)] (approximation coefficient) recursively keeping the output coefficient of each high pass filtering operation [g(n)] (details coefficients) at each stage [24].

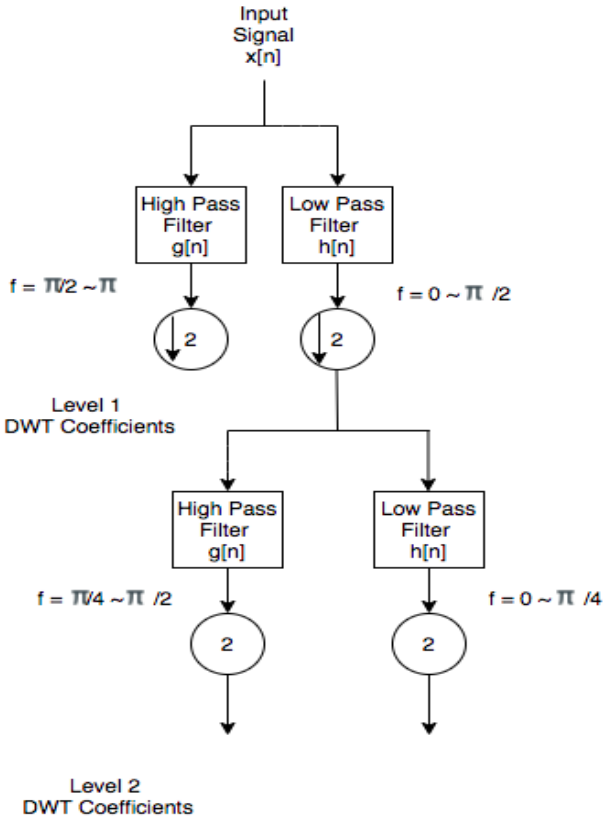

Fig. 1. Shows process of Wavelet decomposition.

Haar: It is one of the most unsophisticated parts of the wavelet family. This is theoretically simple and memory efficient wavelet transform. It uses just two scaling and wavelet function coefficients and decomposes a signal into two sublevels: one is known as an average and the other one is known as difference. This wavelet family looks like a step function and it is non-continuous in nature [14] [17].

Symlet: This wavelet family has the least asymmetry and has densely supported wavelets. The family of Symlet wavelet are the changed form of daubechies wavelets with the increased symmetry. Symlet are symmetric in nature and were proposed by Daubechies (Db) family as amendments [18]. They have comparable properties as Db family. The larger symlets i.e. Sym 12 onwards and have nearly linear phase. These are mostly applicable in smoothing/denoising the applications. They have the identical number of vanishing moments as DbN family [14].

Coiflet: Daubechies family and Coiflets are very similar in a number of ways, but coiflet was constructed with the vanishing moments of wavelet function (phi) and scaling function (psi). The wavelet function has 2N moments and scaling function has 2N-1 moments equal to 0. These functions together have the support 6N-1. The number of vanishing moments is highest in coiflet for a given support width i.e., phi and psi [19]. The wavelet and scaling functions are both normalized by a factor. The scaling function of this family demonstrates the interpolating attributes, that implies excellent approximation of polynomial function at various

resolutions. The symmetrical properties in coiflets are advantageous in signal analysis work due to its linear phase in transfer function. It presents both time and frequency information as essential arrangement [14].

IV. Experimental Methodology

The aim of the experiment is to evaluate the proposed wavelet feature based device identification using its potential as a basis for classifier accuracy. The experimental dataset contains feature shown in Table I.

The data is collected in a monitored environment where we track device activity during data collection. This gives an understanding of the behaviour of the features during the analysis.

To understand the potential of the candidate features in the frequency domain, the hardware features used for evaluation are transformed into wavelet coefficients. These coefficients are then employed as features of the devices for classification. Comparisons are made with existing state-of-the-art methods i.e. logistic Regression and Linear SVC. As shown in Figure 2, the input dataset is provided to the ensemble.

The implementation of three established classification algorithms is examined for device identification, specifically simple logistic regression model, linear SVM and Multivariate Gaussian Distribution classification algorithm [20].

A. Algorithm for the proposed system

The algorithms below introduce the process of generating wavelet coefficients to classification

Algorithm:-

Step 1- Split the data into training and test. This split into the ratio of 80:20 respectively.
Step 2- Apply DWT function to generate wavelet coefficients array from feature set.
Step 3- Using training & test data calculate the accuracy of the predicted labels.
Step 4-Benchmarking Apply classifier. Multivariate Gaussian Distribution [20], Logistic Regression [21] and LinearSVC [22].
Step 5- Repeat Step 1 to Step 5 for other feature sets.

The classification results of the three standard classifiers provides a final device identification result, where the hyper parameters used for SVM is the linear kernel (using this kernel we have only one hyperparameter called cost parameter C) and in case of LR we used liblinear solver and for MVGD the parameters (sigma, mu) are estimated using maximum likelihood.

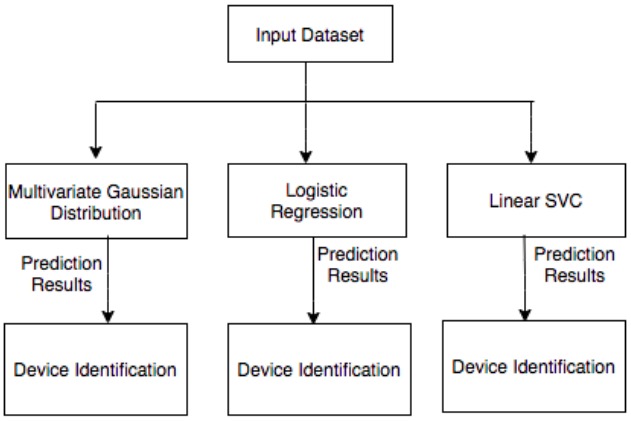

Fig. 2. Device identification using different machine learning models.

B. Classifiers

The fundamental linear classifiers were chosen from sklearn library as they gave the best accuracy amongst the other classifiers based upon the existing features explained in section 3.1.2[23] This work addresses a similarity among three classification techniques assessing which of these methods is best at recognizing and group the devices based on the information collected. In this segment, we present these classification techniques.

1) Simple Logistic Regression Model

Logistic Regression classification model is a well known choice for modeling binary classifications. For this model, the restrictive likelihood of one of the two output classes is assumed to be equivalent to a linear combination of the input features [21] [22].

2) Support Vector Machine Model

We use SVC for support vector machine classification algorithm. The Linear Support Vector Classifier (SVC) technique uses a linear kernel function to do classification and it does as expect well with a large number of samples. If we match it by the SVC model, the Linear SVC has extra parameters such as penalty normalization which applies 'L1' or 'L2' and loss function [22].

3) Multivariate Gaussian Distribution(MVGD)

Any Multivariate Gaussian distribution depicts a vector of several Gaussian distributions in a way that any combination of variables also illustrates the Gaussian distribution. Each of these Gaussians are represented by values derived from the distribution i.e. mean, covariance of the data collected. It identifies the collective distribution of these variables and their mutual probability. Hence the collective effect of the variables is analyzed and probability of each vector is calculated against each distribution [20].

V. Experimental Results

This section presents a discussion of the obtained experimental results with the standard classifiers mentioned above and then comparing the results with raw feature data. K-fold validation is applied on the data where k= 10 and then data is divided into training and test and after analysis our system gives us device identification results based on

classification metrics accuracy percentage, accuracy is defined as the ratio of correct predictions for the test data.

For this experiment we are using wavelet coefficients (Approximation and detail) where we have five hundred samples per device.

Features are selected, based upon high intersample and low intra sample variance amongst features.

The Tables II, III & IV below shows results based upon wavelet feature and raw features for each feature set respectively. For wavelet features we used three different mother wavelets to generate coefficients and we compared results to see which of three wavelet yield best results. For all three feature set Sym6 wavelet give better results than Haar and Coif for device identification. MVGD classifier shows the best accuracy results with Sym6 wavelet.

When compared the results from Table II and Table V, Table II (Sym6 gives highest accuracy results), Table V (Comparing accuracy results from raw features). When compared these two results, classifier with wavelet coefficients give better results than without wavelet transform.

TABLE II. Shows Accuracy Results using Sym6 for all three feature sets.

Sym6			
Classifier	**Accuracy(Approximation)**		
	FS1	***FS2***	***FS3***
MVGD	**92.5%**	**82.7%**	**71.5%**
LR	80.9%	81.9%	51.9%
SVM	91.2%	86.5%	58.1%
	Accuracy(Details)		
	FS1	***FS2***	***FS3***
MVGD	**90.4%**	88.1%	41.8%
LR	83.7%	80.7%	46.9%
SVM	90.3%	**89.5%**	**49.6%**

TABLE III.Shows Accuracy Results using Haar for all three feature sets.

Haar			
Classifier	**Accuracy(Approximate)**		
	FS1	***FS2***	***FS3***
MVGD	90.2%	91.6%	**70.8%**
LR	76.7%	89.1%	40.5%
SVM	**92.7%**	**94%**	42.7%
	Accuracy(Details)		
	FS1	***FS2***	***FS3***
MVGD	70.2%	**93.3%**	**51.7%**
LR	76.8%	90.3%	41.6%
SVM	**77.5%**	90.4%	48.8%

TABLE IV. Shows Accuracy Results using Coif1 for all three feature sets.

Coif1			
Classifier	**Accuracy(Approximate)**		
	FS1	***FS2***	***FS3***
MVGD	89.8%	88.9%	**72%**
LR	88.6%	84.8%	59.8%
SVM	**94.0%**	**90.6%**	58.5%
	Accuracy(Details)		
	FS1	***FS2***	***FS3***
MVGD	89.1%	**85.1%**	**59.8%**
LR	87.9%	90.6%	50.2%
SVM	**92.4%**	92%	49%

TABLE V. Shows Accuracy Results without wavelet transform for all three feature sets.

Without Wavelet Transform			
Classifier	**Accuracy(Without Wavelet Transform)**		
	FS1	***FS2***	***FS3***
MVGD	**89.5%**	**81.6%**	**70.5%**
LR	87%	80.9%	50.8%
SVM	89.3%	86.3%	57.9%

From the research results, we see that Multivariate Gaussian Classifier performs better when compared with other two classifiers in the expectation of identifying devices, particularly using wavelet features. The accuracy results for all three feature sets come out to be better when compared between wavelet features and raw features.

VI. Conclusion

This paper presents a novel wavelet feature based device identification. The device identification technique is compared to standard classifiers. In addition to the wavelet feature based classification, the results are compared to the raw features based classification. Here, we conclude that the device identification using wavelet features yields 92.5% of accuracy in comparison with raw features. Overall wavelet features give better results compared to raw features and Sym6 performs best out of three wavelets.

References

[1] Security requirements for cryptographic modules. (2001). [online] Available at: https://csrc.nist.gov/publications/detail/fips/140/2/final.

[2] CQURE Academy. (2016). Decrypting SID-protected PFX Files Without Having a Password. [online] Available at: https://cqureacademy.com/blog/windows-internals/decrypting-sid-protected-pfx-files-without-password [Accessed 26 Jun. 2021].

[3] Security Research Report on Mercedes-Benz Cars. (n.d.). [online] Available at: https://i.blackhat.com/USA-20/Thursday/us-20-Yan-Security-Research-On-Mercedes-Benz-From-Hardware-To-Car-Control-wp.pdf [Accessed 26 Jun. 2021].

[4] Security Today. (2021). The Next Generation -- Security Today. [online] Available at: https://securitytoday.com/articles/2018/12/01/the-next-generation.aspx [Accessed 26 Jun. 2021].

[5] R. Tahir, H. Hu, D. Gu, K. McDonald-Maier, and G. Howells "Resilience against brute force and rainbow table attacks using strong ICMetrics session key pairs," in Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on, 2013, pp. 1–6.

[6] R. Tahir and K. McDonald-Maier, "Improving Resilience against Node Capture Attacks in Wireless Sensor Networks using ICMetrics," in Emerging Security Technologies (EST), 2012 Third International Conference on, 2012, pp. 127–130.

[7] B. Ye, G. Howells, and M. Haciosman, "Investigation of Properties of ICMetrics in Cloud," in Emerging Security Technologies (EST), 2013 Fourth International Conference on, 2013, pp. 107–108.

[8] A. Hopkins, K. Mcdonald-Maier, and G. Howells, "Device to generate a machine specific identification key." Google Patents, 2013.

[9] www.f-secure.com. (n.d.). Caligula Description | F-Secure Labs. [online] Available at: https://www.f-secure.com/v-descs/calig.shtml [Accessed 5 Jul. 2021].

[10] Y. Kovalchuk, H. Hu, D. Gu, K. McDonald-Maier, D. Newman, S. Kelly, and G. Howells, "Investigation of Properties of ICMetrics Features," in Emerging Security Technologies (EST), 2012 Third International Conference on, 2012, pp. 115–120.

[11] Ghodratnama, S. and Boostani, R. (2015). An Efficient Strategy to Handle Complex Datasets Having Multimodal Distribution. ISCS 2014: Interdisciplinary Symposium on Complex Systems, pp.153–163.

[12] G. Howells, E. Papoutsis, A. Hopkins, and K. McDonald-Maier, "Normalizing Discrete Circuit Features with Statistically Independent values for incorporation with in a highly Secure Encryption System," in Adaptive Hardware and Systems, 2007. AHS 2007. Second NASA/ESA Conference on, 2007, pp. 97–102.

[13] S.Yadav, G. Howells, "Analysis of ICmetric features/technology for wearable devices IOT sensors," in Emerging Security Technologies (EST), 2017 Seventh International Conference on, 2017, pp. 175–178.

[14] Majumdar, Swatilekha. (2013). Comparative Analysis of Coiflet and Daubechies Wavelets Using Global Threshold for Image De-Noising. International Journal of Advances in Engineering and Technology. 6. 2247-2252.
[15] A Wavelet Tour of Signal Processing - 3rd Edition. [online] Available at: https://www.elsevier.com/books/a-wavelet-tour-of-signal-processing/mallat/978-0-12-374370-1 [Accessed 27 Jun. 2021].
[16] Sridhar, S., Rajesh Kumar, P. and Ramanaiah, K.V. (2014). Wavelet Transform Techniques for Image Compression – An Evaluation. International Journal of Image, Graphics and Signal Processing, 6(2), pp.54–67.
[17] Porwik P and Lisowska A. The Haar-wavelet transform in digital image processing: its status and achievements. Mach Graph Vision 2004; 13: 79–98.
[18] J. Kaur and R. Kaur, "Biomedical images denoising using symlet wavelet with wiener filter," 2013.
[19] S. G. Narkhedkar and P. K. Patel, "Recipe of speech compression using coiflet wavelet," in 2014 International Conference on Contemporary Computing and Informatics (IC3I), 2014, pp. 1135–1139.
[20] D. Reynolds, "Gaussian Mixture Models," in Encyclopedia of Biometrics, 2015, pp. 827–832.
[21] A. Singh, N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 1310-1315.
[22] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39.
[23] Scikit-learn.org. (2019). 1. Supervised learning — scikit-learn 0.21.3 documentation. [online] Available at: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning.
[24] al-Qerem, A., Kharbat, F., Nashwan, S., Ashraf, S. and blaou (2020). General model for best feature extraction of EEG using discrete wavelet transform wavelet family and differential evolution. International Journal of Distributed Sensor Networks, 16(3), p.155014772091100.