Chapter 3

EaZy Learning: An Adaptive Variant of Ensemble Learning

Spoof detectors benefit fingerprint authentication systems in terms of increased security and user confidence [69, 70]. Spoof detection becomes a challenging task when spoof fingerprints fabricated using new materials (that were not used in training) or sensed using unknown sensors are introduced to the detectors. Thus, making the fingerprint liveness detection an open set problem [5]. Spoof detectors behave inadequately in the presence of fingerprints generated using fabrication materials that are currently unknown to the detectors [38].

Spoof detectors are majorly categorized into hardware-based and software-based mechanisms [70]. In hardware-based mechanisms, a specialized hardware device monitors the additional life characteristics such as blood pressure, temperature, dry, moist or wet skin. Hardware-based solutions are useful in a static environment, but when the attacker finds a way to crack the system, it becomes difficult to upgrade the hardware device. Also, these devices are unprotected to attacks using new fabrication materials.

Software-based spoof detection mechanisms rely on the representational features of fingerprint images to predict the class labels and are more appropriate for dynamic environments [71]. Various features have been used and proved to be efficient in training the spoof detectors for accurate classification. Still, the spoof detectors have a room for improvement in their performance while testing under cross-dataset and cross-sensor environments. The current state-of-the-art methods suffer from poor generalizability when tested under these environments.

Therefore, the spoof detector must be robust to the changing environment and must perform reasonably well on cross-sensor and cross-dataset settings.

3.0.1 Learning Paradigms for Spoof Fingerprint Detection

Spoof fingerprint detection is an application of biometrics and forensic science where the task is to classify fingerprint images into "live" or "spoof". Ensemble learning has proved to be an adequate solution to this problem, but it does not provide adaptiveness towards the data. We claim that applications like spoof detection require the learning model to be adaptive to the properties intrinsic to the dataset. Therefore, we propose a novel learning scheme, EaZy learning which is midway between eager and lazy learning.

Eager learning compiles the training data greedily and generates a concise hypothesis from the input samples and uses it for decision making. In contrast, lazy learning [11] uses the input samples for decision making. Lazy learning is suited for applications where it is required to have good local approximations. Still, lazy learning requires to store the entire training data and defer the process of prediction until a query appears, which causes high memory consumption and low prediction efficiency. Thus making it challenging to use with practical applications. Lazy learning incurs low computational cost during training but the high cost in responding to the queries.

Therefore, we propose a novel learning model EaZy learning to overcome the challenges in learning paradigms. EaZy learning overcomes the high storage requirements and low prediction efficiency while maintaining good local approximations. The proposed model can be considered as a variant of ensemble learning which considers the properties of data and moves towards the eager or lazy nature of the learning paradigms.

EaZy learning differs from ensemble learning in the way it generates the ensemble and the way it integrates the outputs of the members of the ensemble. One of the major requirements of ensemble learning is to have a pool of diverse base classifiers [72]. We achieve this by performing clustering on the training set and training the base classifiers on each cluster. In that way, we deliver diversity, which results in different generalization capabilities of base classifiers in the ensemble. EaZy learning is a plug-in solution capable of working with various base classifiers on any application.

In general, spoofs sensed using new sensors (i.e., unknown to the learning model) may appear as test instances. Therefore, it is challenging for the spoof detector to keep

(a) Live (b) Ecoflex (c) Gelatin (d) Latex (e) Silgum (f) WoodGlue

Figure 3.1: Visual comparison between live and spoofs created using various spoof materials.

on updating itself. Figure 3.1 shows the visual comparison between live and spoof fingerprints. As can be observed, the human eye can not identify presentation attacks made on biometric systems.

Software-based solutions require extracting the features from the fingerprint images and classifying the image based on the learning model trained over those features [4–6, 27, 31]. In this study, we propose an adaptive model which considers the spoof detection as a binary classification problem. In the past, the study of spoof fingerprint detection has targeted it as an application of closed-set supervised classification. We claim that cross-sensor and cross-dataset performances are of utmost importance, and the research in this field must try to tackle these difficulties. Therefore, we solve this problem by considering it as an open-set problem and make sure that the performance of the spoof detector is minimally compromised. The contributions made in this chapter are as follows:

1. We propose an adaptive learning model EaZy learning which generates an ensemble of diverse base classifiers.

2. We evaluate the performance of various ensemble learning models and EaZy learning for spoof fingerprint detection under cross-sensor and cross-datasets environments.

3. We emphasize on adapting to the properties of data while generating the hypotheses and show robustness against the fingerprints generated using unknown fingerprint sensors.

3.1 EaZy Learning

Our proposed work EaZy Learning uses the same foundation as ensemble learning, i.e., instead of taking one expert's opinion, let several experts discuss and come up with a decision. In Multiple Classifiers Systems (MCSs), our goal usually is to generate a set Π containing n hypotheses that are accurate and diverse [12]. Therefore,

$$\Pi = \{\psi_1, .., \psi_n\} \tag{3.1}$$

where Π is an ensemble of base classifiers ψ_i, and ψ is defined as,

$$\psi : H \times T \rightarrow Y \tag{3.2}$$

where H is a hypothesis which operates on a set T of instances and results in a class label belonging to a set Y. Ideally, ensemble learning is expected to generate a pool of classifiers Π, such that it exploits the unique competencies of each base classifier ψ_i.

The generated hypotheses should be consistent with the subset of data D_{ψ_i} on which they are trained and disjoint from each other, such that:

$$D_{\psi_1} \cup D_{\psi_2} ... \cup D_{\psi_n} = D \tag{3.3}$$

and,

$$D_{\psi_i} \cap D_{\psi_j} = \phi \tag{3.4}$$

EaZy learning satisfies Equations 3.3 and 3.4 by performing clustering on the training data D and training base classifier ψ_i on each D_i. It accommodates the nature of the given data and works well irrespective of its similarity quotient. It is capable of overcoming the drawback of eager learning, i.e., poor local approximations, the drawback of lazy learning, i.e. inefficiency in the classification phase, and the drawback of ensemble learning, i.e. lack of adaptiveness towards the data. To achieve an adaptive midway by maintaining more than one hypothesis, we need to extract the common features of the examples and group them according to the similarity inherently present in them.

As represented in Figure 3.2, we start with passing the training data into the training phase where data is partitioned into n clusters based on the similarity present in the data. Later, we use base learners to generate hypotheses from these clusters. Each cluster c_i yields one hypothesis ψ_i, which is trained only on records belonging to that cluster. The

Algorithm 1: EaZy learning for training a classifier Π.

Input: Training dataset D, validation data V, classifier learning algorithm K, clustering algorithm C

Output: A set of classifiers Π, a set W containing the weights for the classifiers in Π

$\{c_1, c_2, .., c_n\} \leftarrow C(D)$.

Perform clustering on training set D.

for i =1 to n **do**

  $\psi_i \leftarrow K(c_i)$.

  $Acc_i \leftarrow$ Performance of ψ_i on V

end for

$\Pi \leftarrow \{\psi_1, \psi_2, .., \psi_n\}$

$W \leftarrow \{w_1, w_2, .., w_n\}$

Weights determined by Equation 3.5

return Π

return W

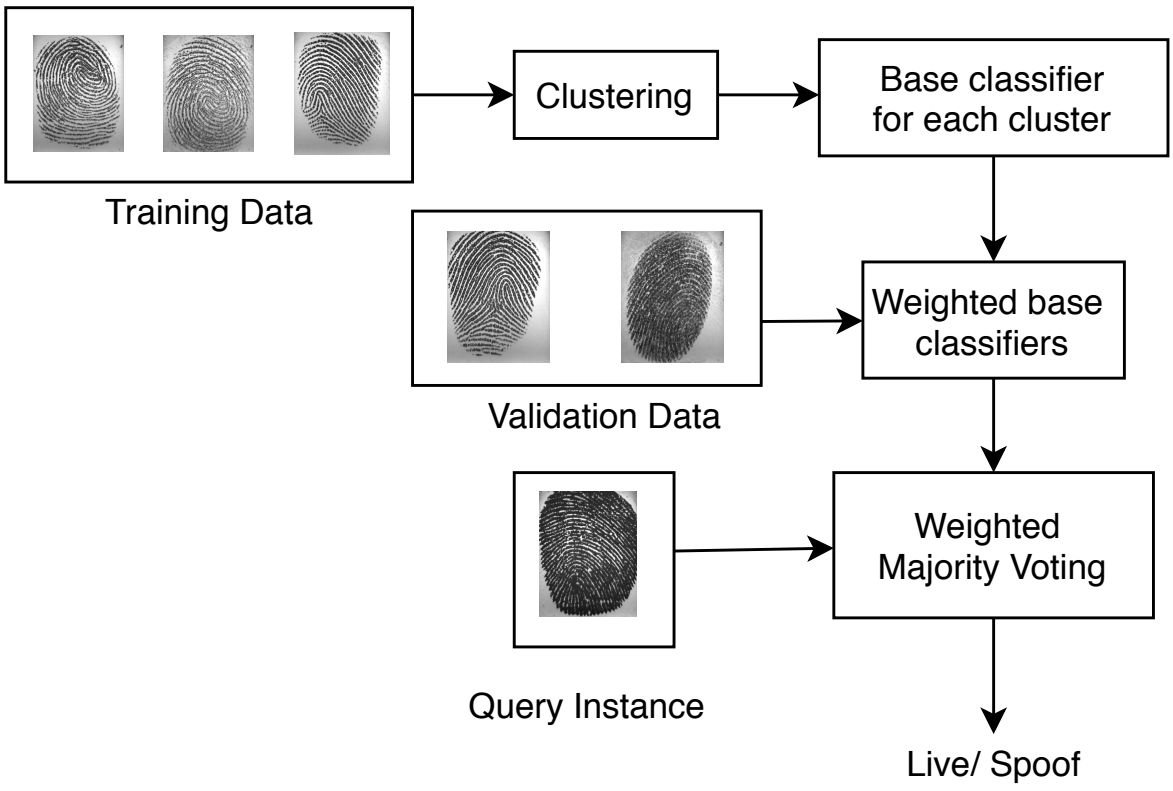

Figure 3.2: Conceptual Model of Adaptive Ensemble Learning [6].

output of the training phase is n hypotheses $(\psi_1, \psi_2, .., \psi_n)$ that are accurate and diverse. The procedure of generating hypotheses for classification is given in Algorithm 1. These hypotheses are given to the testing phase, where a query instance x_q is assigned a discrete class by using weighted majority voting scheme defined by Equation 3.5.

Algorithm 1 is used for training a classifier Π. The inputs to the training phase are a training dataset D, a base classifier K and a clustering algorithm C. We start by generating a validation set V by randomly picking the instances from D. From our experiments, we found that the ideal size of V is 20% of D, but it may be changed depending on the application.

Unlike Ensemble Learning, EaZy learning handles the training data in a different way. As explained in Figure 3.2 and Algorithm 1, EaZy learning generates a number of sub-datasets from the training data after performing clustering on it (Step 1 in Algorithm 1). By performing clustering on the training data, it takes the similarity inherently present in the dataset into consideration. Therefore, we do not require to define the number of classifiers apriori. In the best case, where we encounter a dataset containing all similar records, the proposed method converges itself to the eager learning paradigm and result in one hypothesis only. In the worst case, where we have a dataset containing dissimilar records, the proposed method converges itself to the lazy learning paradigm because the number of hypotheses is close to the number of training examples. On these generated sub-datasets, we use a base learning algorithm to generate the base classifiers or hypotheses (Step 3 in Algorithm 1). These hypotheses are used for predicting the values for query instances.

These classifiers are consistent with the data belonging to the respective cluster. The performance of these classifiers may not be equally good. Therefore, we use V to test the performance of each classifier ψ_i in Π and learn the weightage w_i for each ψ_i as given in Equation 3.5. The validation set V is passed to every ψ_i to check its accuracy A_i. We determine the accuracy A_i as a fraction of the number of instances correctly classified by ψ_i to the total number of instances in V.

A high value of A_i indicates the effectiveness of ψ_i. Therefore its weightage W_i must be directly proportional to A_i. We use the Equation 3.5 to determine the weightage of each ψ_i:

$$W_i = \frac{A_i}{(\sum_i (A_i))} \tag{3.5}$$

We conclude the training phase of our model by generating a set of classifiers H with a set of weights W representing the weightage that should be given to the respective classifier while making a decision.

3.2 Experimental Setup

The experimental setting is designed to demonstrate the working mechanism of EaZy learning and to explore the problem behaviour of spoof fingerprint detection under various environments. With our experiments, we aim to answer the following research questions:

- **RQ1:** How does EaZy learning perform under cross-sensor environment?
- **RQ2:** EaZy learning adapts to the properties of data while generating multiple base classifiers in the ensemble. How is it useful under cross-dataset environment?

3.2.1 Datasets

The description of the datasets used in this study is given in Table 3.1. We use LivDet 2011 [73], LivDet 2013 [74], and LivDet2015 [3] datasets used in fingerprint liveness detection competition held in subsequent years [75]. The goal of this competition is to compare software-based fingerprint liveness detection methodologies and fingerprint systems that are useful in identifying presentation attacks. Each of these datasets consists of live and spoof fingerprint images. These fingerprints are tested on biometric sensors such as Biometrika, DigitalPersona, ItalData, Sagem, CrossMatch etc. For each sensor, we have approximately 1000 fingerprint images belonging to the "live" category and the same number of images belonging to the "spoof" class. We have same number of images in training and testing. Further, the images belonging to the spoof class can be categorized in multiple sub-categories based on the fabrication material used for creating the spoof or fake fingerprint. These materials are gelatin, latex, playdoh, wood glue, silicone etc. The datasets have approximately 200 images, belonging to each of these sub-categories.

Table 3.1: Description of datasets.

Database		Live (Train/Test)	Spoof (Train/Test)
LivDet2011	Biometrika	1000/1000	1000/1000 (ecoflex, gelatin, latex, silgum, wood glue)
	DigitalPersona	1000/1000	1000/1000 (gelatin, latex, playdoh, silicone, wood glue)
	ItalData	1000/1000	1000/1000 (ecoflex, gelatin, latex, silgum, wood glue)
	Sagem	1000/1000	1000/1000 (gelatin, latex, playdoh, silicone, wood glue)
LivDet2013	Biometrika	1000/1000	1000/1000 (ecoflex, gelatin, latex, modasil, wood glue)
	ItalData	1000/1000	1000/1000 (ecoflex, gelatin, latex, modasil, wood glue)
	CrossMatch	1250/1250	1000/1000 (body double, latex, playdoh, wood glue)
LivDet2015	Biometrika	1000/1000	1000/1500 (ecoflex, gelatin, latex, RTV, wood glue)
	DigitalPersona	1000/1000	1000/1500 (ecoflex, gelatin, latex, RTV, wood glue)
	CrossMatch	1510/1500	1446/1448 (body double, ecoflex, gelatin, playdoh, oomoo)

3.2.2 Features

We use ResNet-50 model [76] to extract the features from fingerprint images. ResNet-50 is a deep Residual Network originally designed for object recognition. ResNet-50 has been pre-trained on ImageNet database. By extracting the features using ResNet-50, we utilize transfer learning for spoof fingerprint detection. Due to space constraints, we refrain from discussing the ResNet-50 architecture in detail.

3.2.3 Setup

The motivation for proposing EaZy learning is to be able to generate an ensemble of base classifiers while considering the properties of data. Therefore, we use EM clustering algorithm [77] to generate clusters of training instances. In that way, we get a pool of n disjoint base classifiers at the end of the training phase without defining n apriori. We consider a validation set V which is a hold-out set of the original training data. In this study, we take 20% of the original training data for validation. The remaining 80% data is used as the actual training data. Therefore, we have three separate sets: test, train and validation. The number of clusters obtained in every experiment for EaZy learning is written in parentheses. Experiments were conducted 10 times and the average results are reported.

For applications like spoof fingerprint detection where the false negatives have a huge cost, it is important to report the accuracy of the model along with Attack Presentation

Classification Error Rate (APCER). APCER is defined as:

> *"proportion of attack presentations using the same PAI (Presentation Attack Instruments, e.g., spoof material) species incorrectly classified as bona fide presentations in a specific scenario."*

The conventional ensemble algorithms used for comparison are listed below:

- **Random Sub-Space Method:** Random Sub-Space Method (RSM) [78] is also called attribute bagging because of its nature of bootstrapping the attributes of a dataset to generate new sets. It is a popular ensemble learning approach which attempts to reduce the correlation between estimators in an ensemble by training them on random samples of features instead of the entire feature set. In this study, we use RSM along with SMO [79] as the base classifier.

- **Random Forest:** Random Forest (RF) [80] is a popular ensemble learning algorithm that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

- **Ada-Boost:** Ada-Boost or Adaptive Boosting [81] is also an ensemble learning algorithm that can be used in conjunction with many types of weak learners to boost their performance.

- **Bagging:** Bagging or Bootstrapped Aggregating [82] is an ensemble approach where the dataset is partitioned to create bootstrapped samples. Base classifiers are trained on these samples, and final prediction is made by majority voting. In this study, we use it with SMO base classifier.

In this study, we have considered two experimental settings to explore the behaviour of the proposed model.

1. **Category-1: inter sensor, same material performance evaluation-**
   In Category-1, we evaluate the ability of the spoof detector in the cross-sensor environment. Therefore, we train the model on images acquired from one sensor and test it on images belonging to another sensor. For example, the model is trained on Biometrika 2011 train dataset and tested on ItalData 2011 test dataset. Cross sensor setting evaluates the generalization ability of the model appropriately.

Table 3.2: Performance evaluation of EaZy learning on Category-1.

Dataset	EaZy		RSM(SMO)		Bagging(SMO)		AdaBoost		RF	
	Acc.	APCER	Acc.	APCER	Acc.	APCER	Acc.	APCER	Acc.	APCER
Bio-Ital2011	54.05(13)	0.08	50.5	0.97	51.3	0.91	50.45	0.94	53.5	0.49
Ital-Bio2011	50.5(10)	0.79	51.2	0.41	53.55	0.21	52.55	0.33	47.85	0.94
Sag-Dig2011	50.55(9)	0.83	54.95	0.29	53.9	0.26	54.9	0.25	50.5	0.36
Dig-Sag2011	67.96(16)	0.47	57.35	0.03	56.66	0.02	56.81	0.04	62.01	0.32
Bio-Ital2013	95.4(11)	0.06	83.1	0	82.75	0	72.6	0	74.65	0.48
Ital-Bio2013	57.9(6)	0.83	93.05	0.03	86.75	0.02	91.6	0.02	71.05	0.43
Bio-Dig2015	78.6(7)	0.19	73.16	0.2	73.72	0.16	72.48	0.21	72.8	0.26
Dig-Bio2015	72.16(12)	0.24	62.8	0.56	60	0.61	61.6	0.58	62.68	0.45
Average	65.89	0.44	65.76	0.31	64.83	0.27	64.12	0.3	61.88	0.47

2. **Category-2: inter dataset, same sensor, same material performance evaluation-**

   Category-2 is designed to evaluate the model's capability under cross-dataset environment. These experiments demonstrate the model's robustness against unknown data. Therefore, it is trained on LivDet 2011 and tested on LivDet 2013 and vice-versa.

3.3 Results and Discussion

Next, we answer the research questions raised in Section 3.2 on the basis of our experimental results:

As we mentioned earlier, the problem of fingerprint liveness detection must be projected as an open-set problem, where the test set may contain instances of unknown type, i.e., test-set may have attack presentations generated using various sensors that are not known to the training model. We claim that the proposed model EaZy learning is adaptive to the properties of the dataset. This adaptive nature is useful in cross-sensor and cross-dataset environment. As in real-world scenario, new sensors are used for authentication; a spoof detector needs to detect the new spoofs without requiring to be trained

Table 3.3: Performance evaluation of EaZy learning on Category-2.

Dataset	EaZy		RSM(SMO)		Bagging(SMO)		AdaBoost		RF	
	Acc.	APCER	Acc.	APCER	Acc.	APCER	Acc.	APCER	Acc.	APCER
Bio2011-13	67.1(10)	0.49	59.8	0.56	58.8	0.55	58.6	0.53	51.8	0.72
Bio2013-11	60.7(11)	0.21	52.8	0.94	52.9	0.94	53.35	0.93	53.85	0.91
Ital2011-13	80.55(5)	0.04	67.9	0.63	67.4	0.63	68.65	0.6	63.8	0.59
Ital2013-11	51.15(9)	0.04	51.15	0.98	51.15	0.98	51.4	0.97	54.35	0.89
Bio2013-15	72.68(12)	0.18	63.8	0.46	64.12	0.44	69.36	0.17	40.88	0.98
Bio11-15	46.48(17)	0.03	45.16	0.91	46.56	0.88	48	0.85	40.42	0.93
Bio2015-13	51.5(14)	0.32	49	0.93	49.15	0.88	49.95	0.93	53.85	0.68
Bio2015-11	53.75(7)	0.15	56	0.5	53.85	0.62	52.15	0.49	48.6	0.66
Average	60.49	0.18	55.7	0.74	55.49	0.74	56.43	0.68	50.94	0.79

on those sensors. Table 3.2 and Table 3.3 demonstrate the algorithm's ability to adjust in cross-sensor and cross-dataset environment. In category-1, the performance of EaZy learning is better than its counterparts. The highest accuracy is 95.4% on Bio-Ital 2013 datasets. Similarly, in category-2, EaZy learning outperforms its rivals with the highest accuracy of 80.55% on ItalData 2011-13 datasets.

3.3.1 Discussion

We positioned this study as an adaptive midway between eager and lazy learning. Based on its similarity with ensemble learning (EL), EaZy learning can also be considered as a different variant of Multiple Classifier Systems (MCS) paradigm. Following the motivation, we conducted our experiments under various settings. We focused on the adaptiveness towards the data while deciding the number of base classifiers constituting the ensemble. We emphasized that a spoof detector must learn from the similarity inherently present in the data. Therefore, for a spoof detector to be robust towards the presentation attacks made using unknown biometric sensors, adaptiveness plays an important role. We demonstrated that EaZy learning is best suited for this task. From Table 3.2 and Table 3.3, it is evident that EaZy learning is the best choice for such environments. We

performed Friedman significance test on category-2 APCER values. It is observed that EaZy learning performs significantly better than the rivals where p-value is 0.00112 and ${\chi^2}_r$ value is 18.225. The significance level was set at 0.05. Since p-value is less than the significance lever α, the null hypothesis can be rejected. From these tests we can conclude that the proposed model is significantly better than the rival models under these environments.