Al-Farabi Journal of Engineering Sciences Volume (4), Issue (2) December (2025)

ISSN: 2617-5517 (issn.org)

Al-Farabi Journal of Engineering Sciences

https://iasj.rdd.edu.iq/journals/journal/view/97

مجلة الفارابي للعلوم الهندسية تصدرها جامعة الفارابي

Fraud Detection Based Finger Veins Based Machine Learning Methods Using Support Vector Machine

Ali Shakir Mhaimeed

Sciences & Technology University in Lebanon

Abstract-Finger veins are a unique attribute of the human body. Difficult to trace and understand, since the veins are beneath the skin it is not possible to differentiate which one is larger or more apparent because the individual has more than one finger to evaluate. This article will discuss how fake accounts were identified followed by a look at the preprocessorLDA was utilized for account extraction, and k-fold validation served as a training method. These procedures were conducted on two distinct datasets.. recognizer that performed the best was SVM, and LR had the lowest accuracy of all recognizers..

Keywords: Machine learning, Finger Veins, Imposter, SVM, Liner Regression, One-R.

1. Introduction

Biometrics indicators, which cause the detection of someone "based on their behavioral or physiological properties". The combination of physiological properties that include both hands or fingers, as well as traits on the face, is called iris verification. behavioral traits are characteristics that are acquired or learned. Since finger veins are simple to access, they are highly protected, and they are difficult to duplicate, in contrast to previous eras, more scholars have focused on them. Because of the poor quality of the image and the low porosity "of the Finger's veins, most Finger's veins detection system"s have associated problems with extracting significant features. Poor regulation of the infrared spectrum, insufficient lighting, as well as the tissue that covers the veins in the photographed image can all lead to poor quality images. The four steps of a biometric method that The main steps of a finger vein system include acquiring an image, preprocessing the image, extracting features and matching or verifying. Near infrared light(IR) There are two ways for capturing images: reflected light and transmitted light. In case reflected transmitted A good acquisition process is very important otherwise there will be much work on pre-processing.P reprocessing will have to occur if effective image acquisition does not take place first therefore several contemporary concepts concerning finger vein detection operate effectively with a systematic clear image irrespective changes needed due to quality picture or position finger.V ein'simage should be preprocessed before being acquired so as enhance its fidelity.T wo methods exist for obtaining attributes from an image. The first is to consider the image as a whole, which some articulate feature extraction algorithms in processing images may refer to as an image-level feature extraction method. The second approach removes the shape of the veins in the image, then extracts features from the function of extraction of veins is obtaining simple shapes from veins. Table 1 shows results of the survey for main biometric indicators.

Table1. Survey of main biometric Signs

Biometric Signs	Key Benefit	weakness	Security Level	Cost
Voice	Natural and comfortable	Noisiness	medium	Low
Face	Remote capturing	Brightness Prerequisites	medium	Low
Iris	Guaranteed accuracy	Glasses	excellent	High
Fingerprint	Commonly used	Skin	Good	Low
Finger -veins	Large-safety level	Sickness	excellent	Low

2. Literature Review

Finger vein detection systems have been gained prominence since the beginning of the 21st century. This portion would investigate Early literature that is associated with this research, this would demonstrate the recognition of finger veins via convergent methods, our proposal:

In 2015, He advocated for the use included preprocessed images that improved the quality of the image, detection ROIs and resized images. He also used "Kekre Wavelet Transform method for feature extraction" with 86.3 percent success rate.[15] In 2021 ,they added into their equation an evolved version of kNCN recognizer,this added version is called akNCN (added k-nearest centroid nearest neighbor) was adaptive. The classification rate on the (FV-USM) database is 85.64 percent, which is noteworthy.. In 2019[33] they used discriminant analysis and KNN recognizer to differentiate finger veins .The fidelity features of finger's veins were evaluated summarized using KNN classifier.SDUMLA-HMT dataset accuracies are;KNN=55.84% ,discriminant analysis =92.21%. [5].

3. Finger Veins System Structure Proposed (FVSS)

The proposed (FVSS) employs "a machine-learning algorithm. Finger's veins' images", the definition of the dataset, the pre-processor, the extractor, the K-valator, the classes, and the post-processor are all common features. Figure 1 illustrates the proposed design of the device's architecture (1).

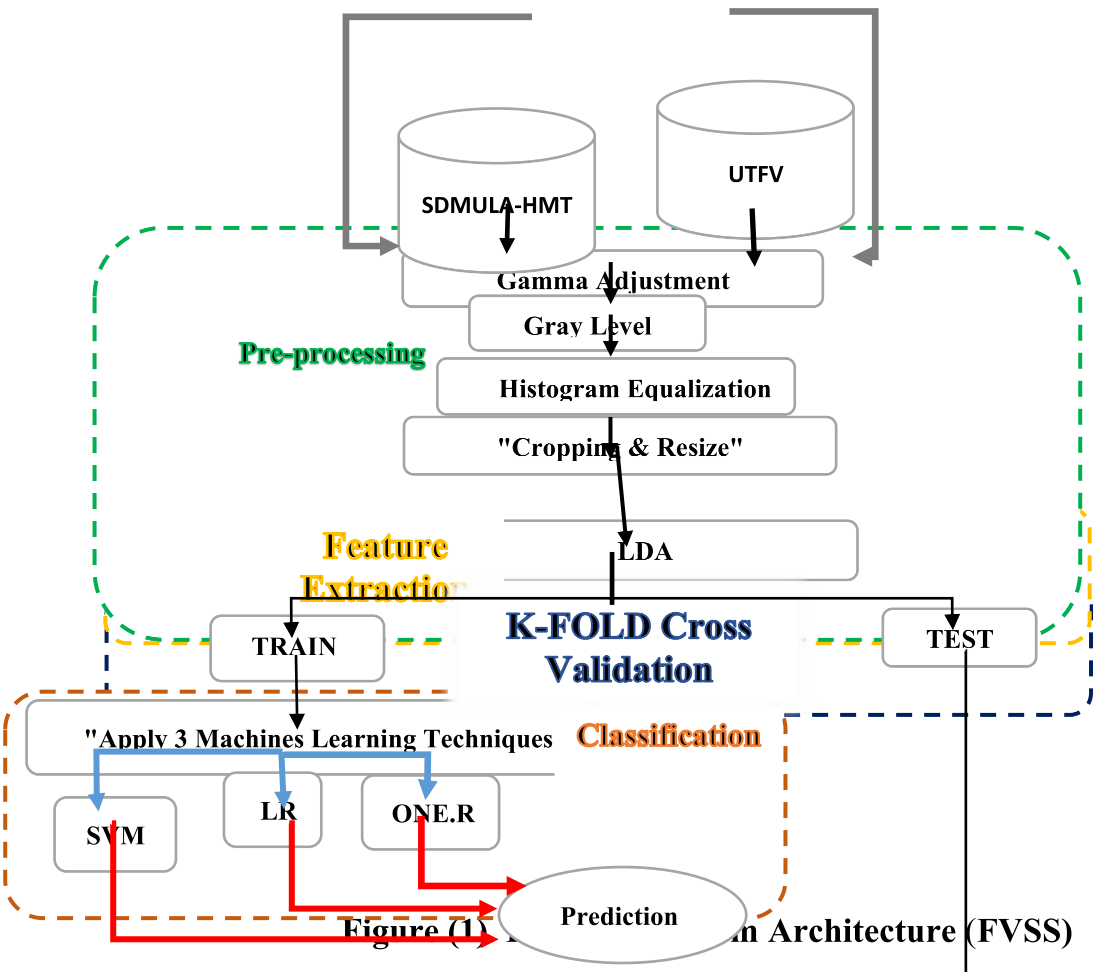
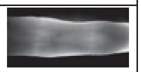


Figure (1) ... n Architecture (FVSS)

3.1. Database Specification

Table 1 shows the specifics of these statistics (2).

"Table (2). Datasets for the System's Finger Veins [1]"

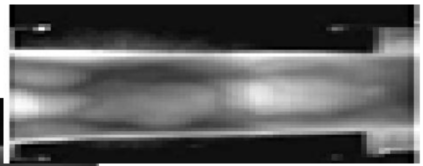
Dataset name	No of images	No of persons	Finger Number per person	Image Number per Finger	Image resolution	Format	Typical Image
SDMULA-HMT	3816	106	3 (Index, ring, middle, of both hands)	6	320x240pxl	.BMP	
UTFV	1440	60	3 (Index, ring, middle, of both hands)	4	672x380pxl	8bit gray scale .PNG	

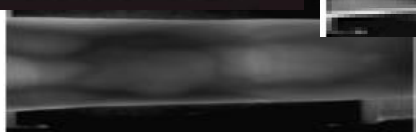
3.2. Preprocessing

The primary benefit of the preprocessing phase is that it facilitates the organization of the data, which leads to an easier process of identification. Both of the aforementioned activities that utilize photographs are considered preprocessing.

3.2.1. "Gamma Correlation for Boost Contrast"

'The fingerprint's image was dark and dim following the capture, so the following step is to make it more bright via gamma correction. A method of regulating pixels that is nonlinear increases the number of effective pixels. This facilitates the minimization of the co-event's involvement in the network. The proposed methodology aims to expose the differences between events and recognize them. This method greatly enhances the quality of the picture by controlling the traditional method of enhancing brilliance in a new way.





"Primary image from (SDMULA-HMT) Database"

"Primary image after **Gamma Correlation"**

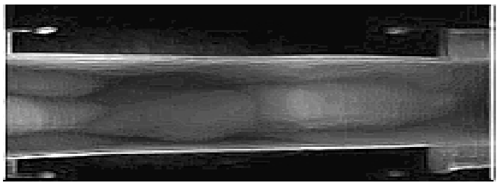
"Primary image after **Gamma Correlation"**

"Primary image from (UTFV) Da

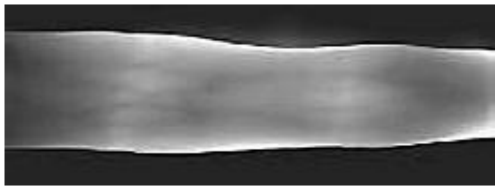
Figure (2). Image Gamma Correlation

3.2.2. RGB to gray level image applying Grayscale Image

Because the color image is marketed as only one channel instead of three separate Red, Green, and Blue channels that convert a Transforming a color image into a grayscale representation results in a reduction of data and a decrease in processing speed. To achieve the conversion from RGB color images to grayscale, a system utilizing weighted averages was implemented. method was preferred because The colors were not distributed with equal weight; green possesses a greater weight than both blue and red.which makes it more significant.



"(SDMULA-HMT) after **Color to Gray level"**

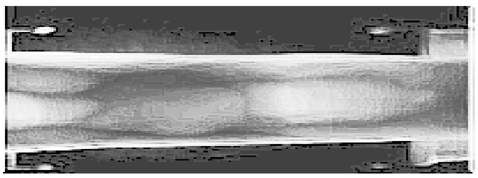


"(UTFV) after **Color to Gray level"**

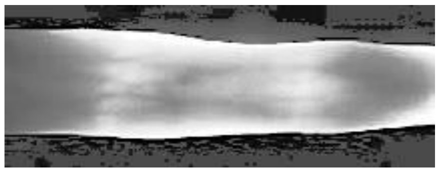
Figure (3). Image Color to Gray level

3.2.3. "Contrast improvement using Histogram Equalization"

'The main objective of histogram equalization is to" standardize the probability distribution of the grey levels in the input image and to redistribute these levels to achieve an image with improved contrast. This technique enhances the overall brightness of the processed image while preserving its original color. Furthermore, it accounts for noise and density-related effects that can result in a degradation of image quality, contributing to the unnatural appearance of the final processed image.



(SDMULA-HMT) after Histogram Equalization

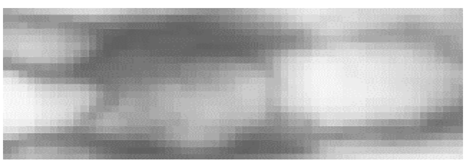


(UTFV) after Histogram Equalization

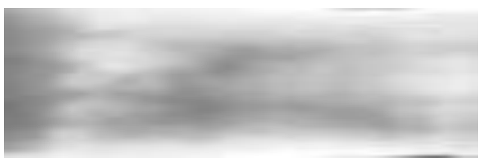
Figure (4). Histogram Equalization for image

3.2.4. "Resizing and Cropping (R&C)"

The most common approaches to picture editing are: (R&C). Both would have been subject to extensive testing because they have the potential to adversely affect the quality of the image. When an image is enlarged, its proportions are changed, which results in a larger file size and, as a result, a higher quality image. Also, cropping involves creating a presentation that is singular, which leads to the loss of pixels.



"SDMULA-HMT after (R&C)"



"(UTFV) after (R&C)"

Figure (5). "Image Resizing and Cropping"

3.3. "Linear Discriminant Analysis For Feature Extraction"

The method of assessment most commonly utilized is Linear Discriminant Analysis (LDA). It employs combinations of variables that are linear in order to most effectively differentiate classes, this results in a linear progression of decisions. The method attempts to transform the data so that the greatest amount of class distinction can be achieved in a low-dimensional space with the greatest possible separation. A single rule that combines linear attributes, axes that define the maximum variation of attributes, and a decrease in the number of attributes can be employed, this is called a dimension.

4. Classifiers of Proposed System

The classification presents a challenge, as a method of feature extraction is essential for evaluating "the overall efficacy of the Finger identification model". Consequently, the subsequent protocols for facilitating classification in machine learning may be utilized.

4.4.1. "Support Vector Machine (SVM)"

The Support Vector Machine (SVM) is capable of categorizing and measuring variables with continuous values. Furthermore, this model effectively addresses the complexities associated with both regression and classification in the realm of data mining. As a machine learning algorithm, the SVM constructs hyperplanes for all class labels within a multi-dimensional space, utilizing margin values. The primary goal of the SVM is to position these hyperplanes as closely as possible to ensure the maximum distinction between classes. A hyperplane is represented by a data case within the specified dataset, which is utilized by support vectors, and margin values are integral to this process. hyperplane.

4.4.2. "One Rule (One-R)"

One-R is a regularity producing decision-making tree employed in data mining. From specific given examples, One-R rules can typically infer simple but very specific classification rules. Even with its simplicity, missing values and numeric attributes do not hinder the applicability of (One-R); hence it has wide potential applications.(One-R) builds a single rule for each training attribute and selects, single rule.

4.4.3. Linear regression (LR)

One-R or One Rule is a regularity-producing decision tree employed in data mining. From a set of instances, One- R typically induces simple but specific rules of classification.(1) Despite its simplicity, potential applications abound since it can also handle missing values and numerical attributes. The (One-R) algorithm generates a single rule for each training attribute and selects the rule with the smallest error measurement as its single rule.

5. Implementation of Proposed Approach

"This method divided into two stages, which are as below":

5.1. Training Phase

First, teach both datasets how to fold in K validation. After that, attach the models to the Finger's veins created by a k fold validation method with k = 6. Gamma will preprocess the dataset, and then convert the images from the RGB color space to the Gray color space, Histogram equalization should be applied to them before cropping or resizing. Classification computers are assigned from functionality presented by LDA. The mixed traits are kept as templates for training phase.

5.2. Testing Phase

The following steps delineate the procedure of the proposed method. As previously indicated, the data that is retained will undergo evaluation at a later stage utilizing the same initial processing steps employed For the training data, The methodology that has been proposed is detailed in the following Algorithm (A).

Algorithm (A), Suggested (FV) implementation
'Input -: Finger veins datasets"
'Output: - Best classifier performance"

Begin

:Import the "dataset (SDMULA-HMT)" and (UTFV). **// Input //**

:: Enhance Contrast through Gamma Correlation Adjustment.

: Convert color images to grayscale levels for the UTFV dataset.

: Enhancement of contrast through the application of Histogram Equalization.

Pre-processing Phase

: "Resizing and Cropping"

: "Dimensionality Reduction" through Linear Discriminant Analysis (LDA)

Feature Extraction Phase

: Randomly rearrange the dataset and divide it into six distinct groups. utilizing Cross-Validation is a method in which (k – 1) subsets are utilized for the purpose of training.

The residual data from the unique sub-instance will be employed as the alidation dataset for testin purposes.

: "Classify instances based 3 Classifiers"

K fold Cross-Validation

0: "Classifiers Evaluation"

1: Optimal performance of the classifier, as indicated by accuracy metrics and error metrics.)

Output

End

6. Proposed System Evaluation

Different metrics are used to assess the behavior of a model during the evaluation of its outputs. The results are based on the volume of training data, the consistency of the files, and the specific type of machine learning algorithm employed. These metrics measure the efficacy of the models..[6]:

- **Accuracy (Acc):** The proportion of instances accurately categorized from the total provided is calculated in the following manner.:

$$\mathbf{Acc} = \frac{\mathbf{aa + bb}}{\mathbf{aa + bb + cc + dd}} \qquad \textbf{(1)}$$

- **Precision (Pre):** For individuals categorized as class x, the proportion of authentic x-class instances is calculated in the following manner.:

$$\mathbf{Pre} = \frac{\mathbf{aa}}{\mathbf{aa + cc}} \qquad \textbf{(2)}$$

- **Recall (Rc):** The ratio of instances designated as class x relative to the total number of instances classified as class x is defined in the following manner.:

$$\mathbf{Rc} = \frac{\mathbf{aa}}{\mathbf{aa + dd}} \qquad \textbf{(3)}$$

- **F- measure(F1):** "precision and recall have a harmonic mean. It's worked out as follows":

$$\mathbf{F_1} = \mathbf{2} * \frac{\mathbf{Pre * Rc}}{\mathbf{Pre + Rc}} \qquad \textbf{(4)}$$

Where

aa = True positives refer to the count of samples that were anticipated to be positive and ultimately resulted in a positive outcome.

cc False positives refer to the count of samples that were anticipated to yield positive results but ultimately produced negative outcomes..

bb = True negatives refer to the quantity of samples that were anticipated to be negative and ultimately were confirmed as negative..

dd= False negatives refer to the quantity of samples that were anticipated to yield negative results but instead resulted in positive outcomes.

- **Error Rate (ErR):** An error fundamentally represents a misclassification, wherein a case is submitted to the classifier, which subsequently misclassifies the case, as illustrated in Equation.. (5) below.:

$$\mathbf{ErR} = \mathbf{1 - accuracy} \qquad \textbf{(5)}$$

- **Specificity(NTR):** The The evaluation of a test's tendency to produce negative results in the absence of the condition is conducted. This occurrence is referred to by several terms, including the false-"positive rate, accuracy, Type I error, error, or null hypothesis".

$$\mathbf{Specificity(NTR)} = \frac{\mathbf{bb}}{\mathbf{bb+cc}}\ \mathbf{100\%}$$

- **KAPPA Cohen's** The kappa coefficient serves to assess the level of agreement between two nominal classifications. When employing Cohen's kappa to evaluate classification agreement, it is assumed that the distances among all nominal categories are uniform. This assumption is reasonable when all nominal categories signify various forms of "presence." "The weighted kappa coefficient is" defined as follows **(6)**

$$:\mathbf{K} = \frac{\mathbf{O-E}}{\mathbf{I-E}}$$

Mean square error (MAE) and RMSE ("Root Mean Square Error"): The accuracy of a recommender system is typically evaluated using these measures, which are computed as demonstrated in Equations (9) and (8). **(7)**

$$\text{"MAE} = \sum |r \wedge n - rn|\ N\ n-1\ /N\text{"} \ldots \quad (8)$$

$$\text{RMSE} = \sqrt{\sum} \text{"}(r \wedge n - rn)\ N\ 2\ n-1\ /N\text{"} \ldots \quad (9)$$

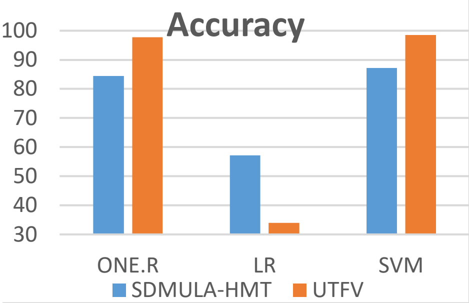
Within this framework, \(r^n \) signifies the anticipated rating, \(r_n \) indicates the actual rating found in the testing dataset, and \(N \) denotes the total count of rating prediction pairs present between the testing data and the prediction outcomes.

7. Experiential Findings

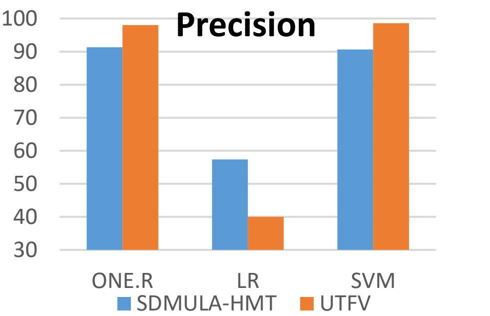
In this experiment, we use classifiers (ORE.R, LR, SVM) on our datasets. Results for the Finger dataset specifically SDMULA-HMT and UTFV are given in Table 1 as referenced from knowledge (2). For SDMULA-HMT and UTFV datasets it is clear from tables (3) figures,(6-14) that out of a The SVM classifier demonstrates a higher overall accuracy percentage, whereas the LR classifier exhibits "the lowest accuracy. In terms of"... of measuring errors results are consistent i.e. out of all the classifiers evaluated LR classifier has a greater error rate while SVM has a least error rate.

"Table (3), Results of Machines Learning" Classifiers

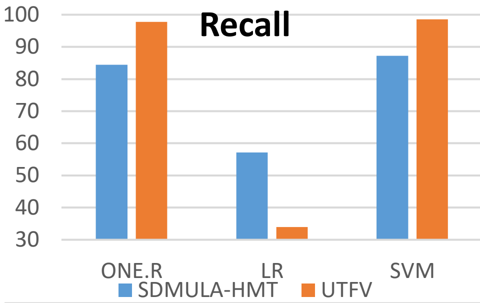
	One.R		**LR**		**SVM**	
	SDMULA-HMT	UTFV	SDMULA-HMT	UTFV	SDMULA-HMT	UTFV
"Total Instances"	3816	1440	3816	1440	3816	1440
"Total correct"	3224	1408	2183	489	3710	١٤19
"Total incorrect"	592	32	1633	951	106	21
"Accuracy"	84.49	97.78	57.21	33.96	87.17	98.55
"Precision"	91.36	97.93	57.37	40.1	90.65	98.62
"Recall"	84.45	97.78	57.26	33.98	87.13	98.56
"F- measure"	85.77	97.72	56.81	34.04	87.19	98.59
"Error rate"	15.51	2.23	42.81	66.04	12.81	1.45
Specificity(NTR)	98.85	98.96	98.59	98.88	98.86	98.96
"KAPPA"	84.33	97.74	56.79	32.83	87.06	98.51
"M.ABS.E"	0.02	0.03	0.009	0.032	0.006	0.003
"RMSE"	0.05	0.14	0.06	0.17	0.03	0.004
Execution Time	6.3 sec	2.34 sec	2.32	2.91 sec	2.2	1.04sec



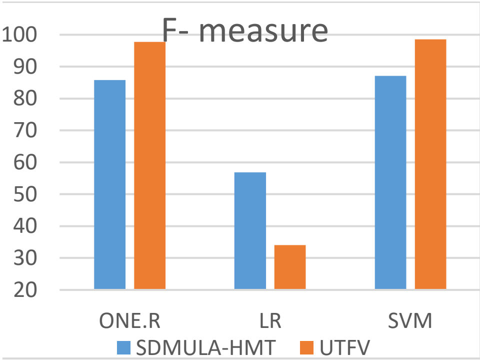
'Figure (6). Acc. for Classifiers"



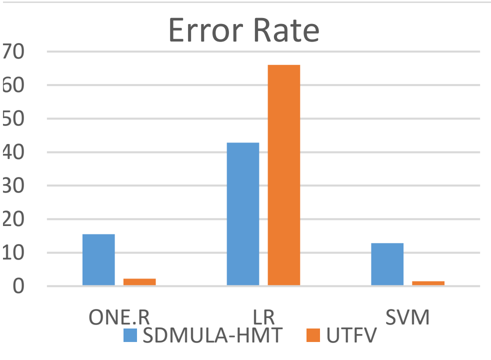
"Figure (7). Pre. for Classifiers"



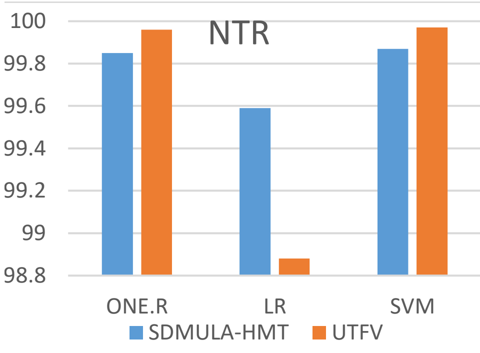
"Figure (8). Rec. for Classifiers"



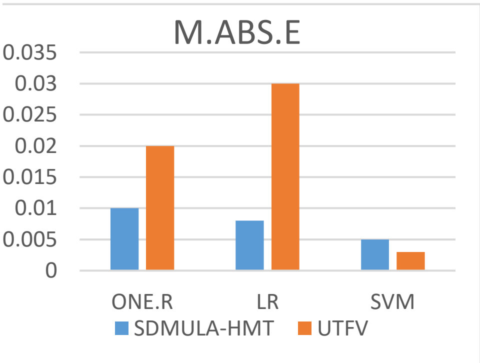
'Figure (9). F-measure for Classifiers"



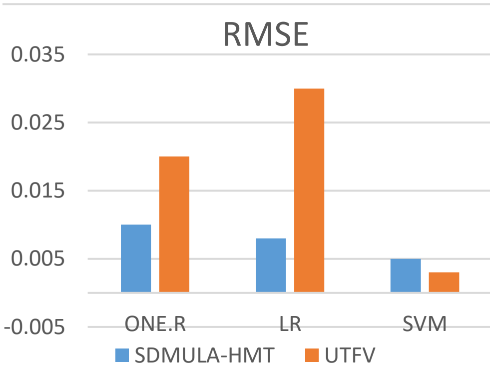
'Figure (10). ER for Classifier"



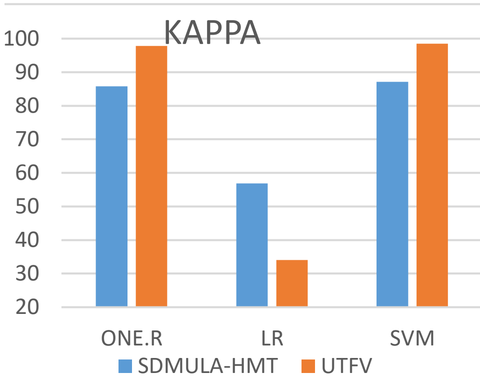
"Figure (11). TNR for Classifier'



'igure (12). M.ABS.E for Classifiers



'igure (13). RMSE for Classifiers



'igure (14). KAPPA for Classifiers

As indicated by the findings presented in Table 3, the Support Vector Machine (SVM) demonstrated a clear advantage "over the other algorithms the One-R algorithm" following closely behind, while the Logistic Regression (LR) algorithm recorded the lowest performance.

8. Results Comparison with Related Work

Table The average precision of the proposed process is assessed using Equation 10, which compares the top two outputs generated by the suggested device algorithms to the pertinent work (4) for the two classifiers, SVM and One-R. (10):

$$\textbf{Average accuracy} = \frac{Acc.of\ dataset1 + Acc.of\ dataset2}{2} \quad (10)$$

Table (4), Results Comparison

Title	**SVM PROPOSED**	(One-R) PROPOSED	References[ ]	References[٤]	References[٥]

Date produced	٢٠٢١	٢٠٢١	2015	2021	2019
Technique used	**SVM**	(One-R)	Euclidean Distance	akNCN	KNN
Accuracy	**93.17%**	91.125 %	86.3%	85.64%	92.21%

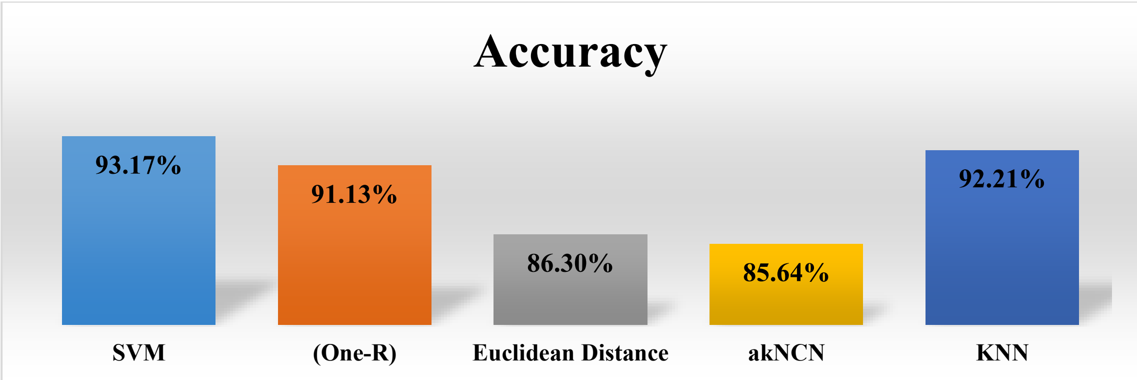


Figure (15). Results Comparison

The results indicate that the suggested approach exceeded the performance of the methods outlined in the Literature Survey regarding The mean accuracy of finger vein detection employing the support vector machine (SVM) algorithm., as illustrated in Table 4.

9. Conclusions

Finger veins are one of the most difficult and complicated biometric traits both to forge and to evade as anti-forensic. Select several different datasets in various formats that will improve and uplift the effectiveness as well as accuracy of the system. Pre-processing images for noise reduction, increased accuracy, rotation, and cropping. Feature extraction is a very crucial process. Using LDA method for feature extraction. In K fold validation testing-training phase k=6 was selected because increasing k up to 10 did not increase accuracy rather reduced time and cost. Different results were generated using machine learning approaches, highest average accuracies found by (SVM) 93.17% while (One-R) took only 91 .125% time therefore we have explored and analyzed the results in all perspectives.

References

[1] Ashraf Tahseen Ali, Hasanen S. Abdullah, and Mohammad N. Fadhil. **"Intelligent system for imposter detection: Asurvey",** Materials Today: Proceedings (2021).

[2] Ashraf Tahseen Ali, Hasanen S. Abdullah, and Mohammad N. Fadhil. **"Finger veins recognition using machine learning techniques",** Materials Today: Proceedings (2021).

[3] Ting, Eiwei, and M. Z. Ibrahim. **"A review of finger vein recognition system",** Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10, no. 1-9 (2018): 167-171.

[4] Rosdi, Bakhtiar Affendi, Nordiana Mukahar, and Ng Tze Han. "**Finger Vein Recognition Using Principle Component Analysis and Adaptive k-Nearest Centroid Neighbor Classifier",** International Journal of Integrated Engineering 13, no. 1 (2021): 177-187.

[5] Khanam, Ruqaiya, Ramsha Khan, and Rajeev Ranjan. **“Analysis of finger vein feature extraction and recognition using DA and KNN methods”,** In 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 477-483. IEEE, 2019.

[6] Ashraf Tahseen Ali, Hasanen S. Abdullah, and Mohammad N. Fadhil. **“Voice recognition system using machine learning techniques”,** Materials Today: Proceedings (2021).

[7] A. Pérez Vega, C. M. Travieso and J. B. Alonso, "Biometric personal identification system based on the patterns created by the finger's veins", the 3rd IEEE International Conference on Bio-inspired Intelligence, p. 65-70, 2014.

[8] K. Wang, X. Yang, Z. Tian and T. Du, "The Finger Vein Recognition Based on Curvelet", Proceedings of the 33rd Chinese Control Conference, pp. 4706-4711, 2014. Available: 10.1109/ChiCC.2014.6895733.

[9] E. Ting and M. Z.Ibrahim, "A review of finger vein recognition system", Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-9, pp. 167-171, 2018.

[10] Ali, A. T., Abdullah, H. S., & Fadhil, M. N. (2021). Impostor Detection Based Finger Veins Applying Machine Learning Methods. IRAQI JOURNAL OF COMPUTERS, COMMUNICATIONS, CONTROL AND SYSTEMS ENGINEERING, 21, 98-111.