World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:5, 2021

The Reproducibility and Repeatability of Modified Likelihood Ratio for Forensics Handwriting Examination

O. Abiodun Adeyinka, B. Adeyemo Adesesan

Abstract**—**The forensic use of handwriting depends on the analysis, comparison, and evaluation decisions made by forensic document examiners. When using biometric technology in forensic applications, it is necessary to compute Likelihood Ratio (LR) for quantifying strength of evidence under two competing hypotheses, namely the prosecution and the defense hypotheses wherein a set of assumptions and methods for a given data set will be made. It is therefore important to know how repeatable and reproducible our estimated LR is. This paper evaluated the accuracy and reproducibility of examiners' decisions. Confidence interval for the estimated LR were presented so as not get an incorrect estimate that will be used to deliver wrong judgment in the court of Law. The estimate of LR is fundamentally a Bayesian concept and we used two LR estimators, namely Logistic Regression (LoR) and Kernel Density Estimator (KDE) for this paper. The repeatability evaluation was carried out by retesting the initial experiment after an interval of six months to observe whether examiners would repeat their decisions for the estimated LR. The experimental results, which are based on handwriting dataset, show that LR has different confidence intervals which therefore implies that LR cannot be estimated with the same certainty everywhere. Though the LoR performed better than the KDE when tested using the same dataset, the two LR estimators investigated showed a consistent region in which LR value can be estimated confidently. These two findings advance our understanding of LR when used in computing the strength of evidence in handwriting using forensics.

Keywords**—**Logistic Regression LoR, Kernel Density Estimator KDE, Handwriting, Confidence Interval, Repeatability, Reproducibility.

I. Introduction

EVIDENCE needs to be weighed more objectively using the biometric technology for the purpose of deciding whether a particular person has committed a crime. Biometric technologies that use face, gait and handwriting biometrics are beginning to be accepted [1]-[3]. Handwriting focuses on various strokes and their relation to personality of an individual. Forensic investigators use handwriting pattern to determine personality traits of an individual. In fact, different agencies use graphology for job application, recruitment procedure, career guidance and child behavior and development [10]. Using handwriting to identify a person is of great importance to justice and law enforcement systems.

Abiodun Adeyinka O. is with the Department of Computer Science & Information Technology, Bowen University, Iwo, Nigeria (e-mail: adeyinka.abiodun@bowen.edu.ng).

Adeyemo Adesesan B. is with the Department of Computer Science, University of Ibadan, Ibadan, Nigeria (e-mail: sesanadeyemo@gmail.com).

The LR or probability proportion paradigm has been studied as method for measuring the quality of proof for a variety of forensic evidence types. An LR is not testing a pair of competing propositions but rather seeking a measure of the relative support of a particular piece of evidence for the validity of one proposition H_p vs. another proposition H_d. The interpretation of a LR equal to say V is that the evidence is V times more likely to have been observed if H_p is true than if H_d is true [8] In theory, LR is an estimate of confidence. By and by, since data are used to estimate the confidence value (i.e., the probability proportion), the estimand can be influenced by the manner in which the data have been sampled so there is a need to test for the repeatability and reproducibility of the estimates probability proportion.

A few inspectors choose to utilize same data and parameters for investigation but there have been a few investigations showing that inspector choices are not generally in understanding [4]-[7]. Earlier work on repeatability [28]-[30] has exhibited that changed choices happen under both biasing and non-biasing conditions. Our study evaluated the accuracy, repeatability and reproducibility of examiners' decisions for document examination. In this paper, we conducted a retest of a period of six months to determine if we would get a repeat of the results gotten earlier and used two different estimator to also determine if these two estimators are reproducible under the same data set. The findings of this investigation fortify the comprehension of forensic document examiners' decisions, contributing to the scientific basis for handwriting examination. Forensic science community needs to clarify the value of forensic evidence with respect to legal questions of admissibility; by assisting with recognizing where to center preparing, confirmation, and normalization; and by providing data to assist agencies in managing finite resources and improving techniques to guarantee the nature of results.

The probability of repeatability and reproducibility of examination by different examiners depends on so many factors. The type of examination performed is one factor [9], [11]-[14]. Repeatability can vary from examiner to examiner, while reproducibility can vary by subpopulation (for example similar training). It is expected that same pre-processing, dataset and examination is used when testing for repeatability and reproducibility meaning that the quality and quantity of corresponding information present in a pair of images is either very high or very low in both cases and examinations. Exact equal rates of agreement are not expected for individualization decisions under reproducibility but it is expected that the two operators or person carrying out the same procedures to test for reproducibility should follow the same definition for agreement, same questions under investigations, same parameters to estimate LR and in summary same rules given that we want to test for the reproducibility using same procedures but a different operator/persons.

Little empirical research on the repeatability and reproducibility of decisions by latent print and face examiners have been carried out so far. Examiners usually agree from what has been published, but are not entirely consistent (Each examiner carries out their own examination based on their level of expertise and experience); so the data being used to estimate the denominator, parameters involved in estimation, question at examination amongst other things vary from one examiner to the other, which is a relatively important factor for the inconsistency.

Reproducibility implies the replicability of result if an estimation is taken by someone else/analyst; though repeatability alludes to the replicability of aftereffect of a similar individual or analyst in another investigator (occurring after a specific time slip by as for the principal investigator) utilizing a similar estimation conditions, i.e., a similar technique, same analyst, same estimating framework, same working condition, same dataset and same area. We might want to discover the comparability of these ideas in measuring LR.

Reproducibility: One approach to evaluate the reproducibility of LR is to think about the accompanying situation. We made use of two LR estimators, for example, LoR and KDE; we need to know whether they would give reliable LR estimates or not. On the off chance that two free estimators give predictable values of LR, at that point the specific LR ought to be more dependable, in light of the fact that the worth is reproducible utilizing two distinct calculations.

Repeatability: The subject of repeatability in our setting could be addressed utilizing time span. Utilizing this system, the equivalent exploratory conditions would be utilized yet with one exemption; the time wherein they are completed changes. In this paper, a time of a half year was given between the main investigation and the subsequent examination. In the event that a specific LR is repeatable, we would expect the value to remain the same within acceptable confidence intervals despite progress in the time span.

II. Methodology

A. Introduction

Some recent work on repeatability and reproducibility utilising using the latent fingerprint and face biometrics were accounted for in [17], [16] while [15] presents examination of three algorithms in evaluating LR, in particular LoR, KDE, and Pull-adjacent Violator algorithms. Their variation in values shows that LR estimates can be inconsistent anywhere. Reference [18] made a clear argument of how they believe evidential value should be presented in a court of law. They contended that it ought to be presented in a single worth value gotten from Bayesian Factor (BF) instead of an articulation dependent on a dispersion over a range of values.

B. Bayesian Factor

A Bayesian factor (BF) is the ratio of the likelihood of a set of data under two different models. These two models might take the form of a null hypothesis (H_0) that a parameter is zero, and an alternate hypothesis (H_1) that the parameter is not zero, although these are not the only types of models that might be compared.

Just as in a frequentist analysis, a Bayes factor comparison of a null hypothesis and an alternate hypothesis considers how probable a particular set of observations would be if the null hypothesis were true, $P(Data|H_0)$. However, a Bayes factor analysis also considers how probable the observations would be if the alternate hypothesis were true. The model under which the data would be more likely is the one whose credibility is improved by the observation of this data.

$$BF = \frac{P(Data|H_1)}{P(Data|H_0)} \tag{1}$$

A question might arise if the probability of a particular data are to be calculated if the alternate hypothesis was true, given that the alternate hypothesis is generally just a vague statement that the true effect size is something other than zero. The answer is that we cannot leave the alternate hypothesis so vaguely specified. Instead, we have to specify a conditional prior probability distribution for the parameter under the alternate hypothesis, which specifies which values of the parameter would be more or less probable if we knew the null hypothesis was false. Some of our prior probability was placed on the null hypothesis being true, and spread the remainder of the prior probability over a range of values.

A Bayes factor is not itself a statement about the posterior probability that a particular hypothesis is correct. However, the Bayes factor comparing a null hypothesis and an alternate hypothesis is the crucial link between the prior odds that the alternate hypothesis is correct and the posterior odds that it is the one correct, taking into account the data observed. If we multiply the prior odds by the Bayes factor BF, the result is the posterior odds.

$$\text{Posterior Odds} = \frac{P(H1|Data)}{P(H0|Data)} = \frac{P(H1)}{P(H0)} \times \frac{P(Data|H1)}{P(Data|H0)} \tag{2}$$

In BF or we can say LR, two contending theories are considered in registering the quality of proof, to be specific:

- Prosecution theory, H_0, bolsters a case that the gathered proof has a place with the suspect, and
- Defense theory, H_1, bolsters a case that the proof has a place with another person.

Given a bit of proof, E, the Bayes decide proposes that the decision ought to be made dependent on the derivation of the back likelihood $P(H_k|E) \propto P(E|H_k)P(H_k)$ for $k \in \{0, 1\}$, which is an outcome of the item rule (the documentation of which is to be additionally clarified in the Methodology section).

Computing the proportion somewhere in the range of $P(H_0|E)$ and $P(H_1|E)$,

Posterior probability ration = LR × prior probability ratio

$$\frac{\Pr(H|E,I)}{\Pr(H_d|E,I)} = \frac{g(E|H_p,I)}{g(E|H_d,I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \quad (3)$$

This is a mathematical formula that represents the belief of a suspect being guilty or innocent as evidence in legal process presented in [19]. This detailing makes forensic expert to weigh in his/her commitment expressly in the form of LR which is a method for measuring the quality of proof and it depends on the criminological assessment on the proof as opposed to the earlier odd proportion which is the jury's transmit. Crime scene serves a great purpose in the context of forensic evidence evaluation for the court of law. For this paper, confirmations were not just constrained to crime scene alone as reported in [27].

Choice is made by the appointed authority; and not legal professionals in a scientific situation unlike the typical utilization of a biometrics framework, when a biometric test from a suspect is contrasted with a bit of biometric proof gathered from the wrongdoing scene, the resultant closeness score is not adequate to be introduced in the courtroom. While acknowledge or reject choice can be made in a commonplace biometric confirmation framework, such a prescriptive methodology is not appropriate for forensic applications [21]-[22].

We only work on the LR and leave the decision making process to the court of law in this study.

C. KDE

The KDE approach directly assesses the LR p(E|Hk) utilizing the KDE calculation, which is a non-parametric strategy. This methodology is reasonable given adequate an example size. KDE places a portion on every information point with the goal that the probability assessed on a given area is a whole of the possibilities of the part work characterized by all the preparation tests. For our motivation in this paper, it does the trick to formalize KDE as an estimator of the structure:

$$P(E|H_k) \cong f(E|Y^k) \quad (4)$$

which is dependent on the training score set Y^k. The approximated Log-likelihood Ratio (LLR) € by KDE is therefore,

$$LLR_{KDE}(E) = \log\frac{f(E|Y^0)}{f(E|Y^1)} \quad (5)$$

Due to the utilization of KDE, the estimate of likelihood on areas where test data are scanty, e.g., at the outrageous tails of the conveyance, can be extremely off base. Hence, another way to deal with displaying LLR (E) is by utilizing LoR.

D. LoR

LoR is a commonly used pattern recognition technique for many problems including fusion and calibration [23]-[26], [15]. In correlation, KDE can be seen as a non-straight change of the crude biometric matcher yield. With the end goal of ensuing conversation, we will digest the LLR evaluated by LR using the following equation:

$$LLR\ LoR\ (E) = fLR(E|Y_0, Y_1) \quad (6)$$

Y_0 and Y_1 are score set prepared by LR using (6). In the wake of preparing, we just need to keep w_1 and w_0 in light of the fact that these two boundaries are together and required to compute LLR. It ought to be forewarned that LR may incidentally display the earlier likelihood of the preparation information. This circumstance is specific intense with lopsided preparing tests, i.e., $|Y_0| \ll |Y_1|$. This can be moderated during the improvement procedure by guaranteeing that each example in Y_0 has a related weight commitment of $1/|Y_k|$ for the two informational collections $k \in \{0, 1\}$.

III. Results

Our paper used two LR estimators KDE and LoR to estimate our modified LR to see if the modified LR is reproducible and repeatable. The two estimators KDE and LoR were used to estimate our modified LR using the same dataset. Fig. 1 shows the results using the LoR estimator while Fig. 2 shows results using KDE with both estimators using the same dataset. For case 1 under the LoR H_p supported (OR > 0.5) 99.4% while for the KDE estimator, H_p supported (LR > 1.00) 97.6%. LoR was against H_p (OR < 0.5) i.e., RMEP (rate of misleading evidence against prosecutor) or disagreement for 0.6% while KDE was against H_p (LR < 1.00) 2.4%. For case 10, under the LoR H_p supported (OR > 0.5) 97.01% while KDE estimator H_p supported (LR > 1.00) 89.22% RMEP under LoR gave H_p (OR < 0.5) 2.99% while that of KDE gave H_p (LR < 1.00) 10.78%. The results of each suspect with other 229 potential suspects in the pool of database were presented in cases as each suspected was tested against every other suspect for our denominator generator because one-on-one exhaustive mapping was used for this with all cases having the maximum of 0% inconclusiveness and both estimators stating the LCI and UCI.

Fig. 3 presents the success and disagreement rate of the LoR method. As shown, the method success rate is high and has minimized (< 40%) disagreement rate in scenarios where disagreement exist in any case file considered. The blue line (success rate) forms above 60% and the red line (disagreement rate) forms below 40%. Thus, out of 230 cases (with each cases consisting of 3 documents or evidences) considered for this study, disagreement only ensued in few cases.

Unlike the LoR method, the KDE method has somewhat higher disagreement rate compared to the LoR method. However, the success rate is higher and the disagreement rate descends, as shown in Fig. 4. In order to have a clear picture of this result, we plot the success rate of the two methods in Fig. 5.

			LR		
	Support Hp (OR > 0.5)	Against Hp (OR < 0.5)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
Case 1	99.4	0.6	98.23219522	100	0.6
Case 2	99.4	0.6	98.23219522	100	0.6
Case 3	98.2	1.8	96.1895493	100	1.8
Case 4	98.8	1.2	97.15346667	100	1.2
Case 5	99.4	0.6	98.23219522	100	0.6
Case 6	89.02	10.96	72.8851066	85.1948934	10.96
Case 7	94.61	5.39	91.19520738	98.02479262	5.39
Case 8	88.44	11.56	72.22137167	84.65862833	11.56
Case 9	96.41	3.59	93.59674113	99.22325887	3.59
Case 10	97.01	2.99	94.43459653	99.58540347	2.99
Case 11	87.43	12.57	82.41698149	92.44301851	12.57
Case 12	84.45	14.55	68.94186129	81.95813871	14.55

Fig. 1 Results obtained from the estimation of the modified LR using LoR Estimator

		KDE			
Case	Support Hp (LR >1.00)	Against Hp (LR <1.00)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
1	97.6	2.4	95.28563443	99.91436557	2.4
2	95.21	4.79	91.98068455	98.43931545	4.79
3	94.61	5.39	91.19520738	98.02479262	5.39
4	94.01	5.99	90.42159157	97.59840843	5.99
5	92.82	7.19	88.91351053	96.72648947	7.19
6	92.22	7.78	88.16954522	96.27045478	7.78
7	91.62	8.38	87.429956	95.810044	8.38
8	91.02	8.98	86.69677325	95.34322675	8.98
9	90.42	9.58	85.9694218	94.8705782	9.58
10	89.22	10.78	84.53033392	93.90966608	10.78
11	88.62	11.38	83.81782	93.42218	11.38
12	86.83	13.17	81.71637097	91.94362903	13.17

Fig. 2 Results obtained from the estimation of the modified LR using KDE

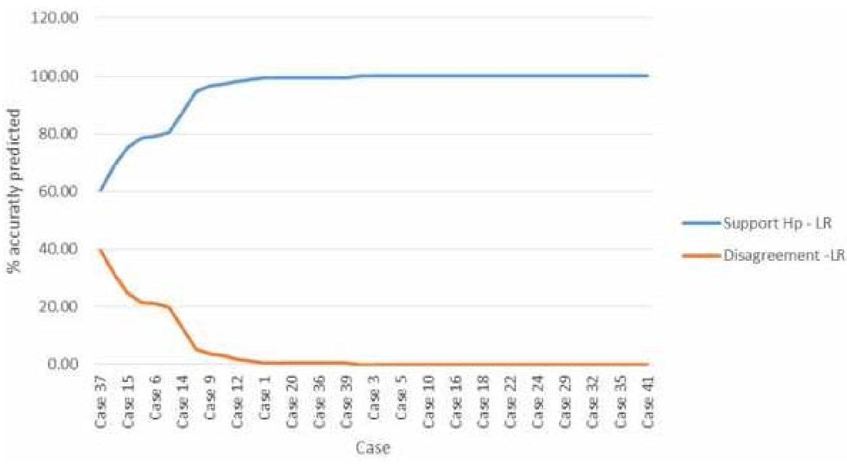

Fig. 3 LoR (Rate of misleading/disagreement in red and success rate in blue line)

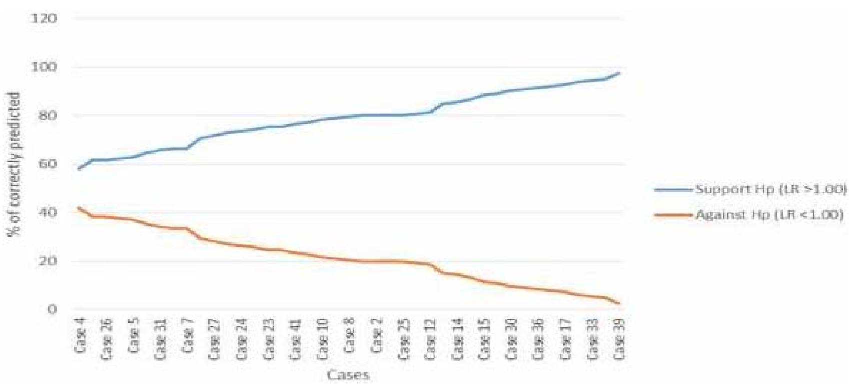

Fig. 4 KDE (Rate of misleading/disagreement in red and success rate in blue line)

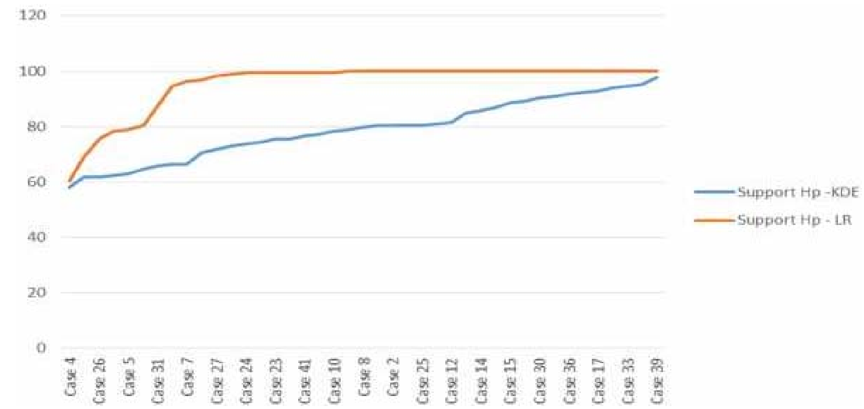

Fig. 5 Success rate for the two estimators

		LR			
Case	Support Hp (OR > 0.5)	Against Hp (OR < 0.5)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
1	100	0	100	100	0
2	100	0	100	100	0
3	100	0	100	100	0
4	99.4	0.6	98.23219522	100.5678048	0.6
5	99.4	0.6	98.23219522	100.5678048	0.6
6	99.4	0.6	98.23219522	100.5678048	0.6
7	99.4	0.6	98.23219522	100.5678048	0.6
8	98.8	1.2	97.15346667	100.4465333	1.2
9	98.2	1.8	96.1895493	100.2104507	1.8
10	97.01	2.99	94.43459653	99.58540347	2.99
11	96.41	3.59	93.59674113	99.22325887	3.59
12	94.61	5.39	91.19520738	98.02479262	5.39

Fig. 6 Results showing estimated LR for our research using the LoR estimator

			LR		
	Support Hp (OR > 0.5)	Against Hp (OR < 0.5)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
Case 1	99.4	0.6	98.23219522	100	0.6
Case 2	99.4	0.6	98.23219522	100	0.6
Case 3	98.2	1.8	96.1895493	100	1.8
Case 4	98.8	1.2	97.15346667	100	1.2
Case 5	99.4	0.6	98.23219522	100	0.6
Case 6	89.02	10.96	72.8851066	85.1948934	10.96
Case 7	94.61	5.39	91.19520738	98.02479262	5.39
Case 8	88.44	11.56	72.22137167	84.65862833	11.56
Case 9	96.41	3.59	93.59674113	99.22325887	3.59
Case 10	97.01	2.99	94.43459653	99.58540347	2.99
Case 11	87.43	12.57	82.41698149	92.44301851	12.57
Case 12	84.45	14.55	68.94186129	81.95813871	14.55

Fig. 7 Results showing estimated LR for our research using the LoR estimator at a different time interval

		KDE			
Case	Support Hp (LR >1.00)	Against Hp (LR <1.00)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
1	97.6	2.4	95.28563443	99.91436557	2.4
2	95.21	4.79	91.98068455	98.43931545	4.79
3	94.61	5.39	91.19520738	98.02479262	5.39
4	94.01	5.99	90.42159157	97.59840843	5.99
5	92.82	7.19	88.91351053	96.72648947	7.19
6	92.22	7.78	88.16954522	96.27045478	7.78
7	91.62	8.38	87.429956	95.810044	8.38
8	91.02	8.98	86.69677325	95.34322675	8.98
9	90.42	9.58	85.9694218	94.8705782	9.58
10	89.22	10.78	84.53033392	93.90966608	10.78
11	88.62	11.38	83.81782	93.42218	11.38
12	86.83	13.17	81.71637097	91.94362903	13.17

Fig. 8 Results showing estimated LR for our research using the KDE

		KDE			
Case	Support Hp (LR>1.00)	Against Hp (LR<1.00)	LCI (95%) support Hp	UCI (95%) support Hp	Disagreement
1	96.67	3.33	92.3875492	98.8226775	3.33
2	97.62	2.38	95.2986641	99.8642764	2.38
3	96.43	3.57	94.8627813	98.274975	3.57
4	95.72	4.28	91.1953792	98.2478917	4.28
5	94.12	5.88	90.387219	97.4729601	5.88
6	92.86	7.14	89.3928751	95.3975507	7.14
7	91.67	8.33	87.4299568	95.2785462	8.33
8	90.91	9.09	88.1695442	94.7539819	9.09
9	88.89	11.11	85.872491	95.287401	11.11
10	87.5	12.5	83.2789521	93.2981738	12.5
11	87.5	12.5	82.9734618	94.6724914	12.5
12	85.71	14.29	81.71637097	90.9436293	14.29

Fig. 9 Results showing estimated LR for our research using the KDE at a different time interval

Fig. 5 shows the success rate for the two estimators used i.e., KDE and the LoR. Both estimators were repeatable and reproducible when tested using our developed model of a modified LR, though the LoR performed better than the KDE.

Figs. 6 and 7 show the results of the investigation using the LoR estimator but with an interval of six month. The same procedure, same experiments, same datasets, same operator were used and roughly same results were gotten. Case 1 for Fig. 6 H_p supports (LR > 0.5) for 100% but for the same case in Fig. 7 (which is at a different time interval) H_p supports (LR > 0.5) for 99.4%, for case 8 under Fig. 6 H_p supports (LR > 0.5) 98.8% while under Fig. 7 case 8 H_p supports (LR > 0.5) 88.44%. LR was against H_p (LR< 0.5) 0% in case 1 under Fig. 6 and 0.6% in case 1 under Fig. 7. For case 8 under both figures, LR was against Hp (LR< 0.5) 1.2% and 11.56% under Figs. 6 and 7 respectively which shows that our system under the LoR estimator is repeatable.

For the KDE, Case 1 under Fig. 8 Hp supports (LR > 1.00) for 97.6% for the same case under Fig. 9 (which is at a different time interval) supports Hp (LR > 1.00) for 96.67%, for case 8 under Fig. 8 H_p supports (LR > 1.00) 91.02% while under Fig. 9 case 8. H_p supports (LR > 1.00) 90.09%. LR was against H_p (LR < 1.00) 2.4% in case 1 under Fig. 8 and 3.33% in case 1 under Fig. 9. For case 8 under both figures, LR was against H_p (LR < 1.00) 8.98% and 9.09% under Figs. 8 and 9 respectively which shows that our system under the KDE is also repeatable.

IV. Conclusion

The test for the repeatability and reproducibility of the modified LR was carried out in this research under a given time interval. It is evident that the repeatability for face and fingerprint biometrics might not be repeatable (for a given time interval). The repeatability and reproducibility of handwriting was also tested (for a giv7en time interval) on the modified LR. It was concluded that though LoR estimator performed better than the KDE in the research carried, both estimator were repeatable and reproduced the same results when test on the same dataset, making our model of the modified LR repeatable and reproducible.

References

[1] R. Hasting. From grainy cctv to a positive id: Recognizing the benefits" of surveillance. The Independent, 2013.

[2] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. Journal of Forensic Sciences, 56(4):882–889, 2011.

[3] P. K. Larsen, E. B. Simonsen, and N. Lynnerup. Gait analysis in forensic medicine*. Journal of Forensic Sciences, 53(5):1149–1153, 2008.

[4] I.E. Evett, R.L. Williams. A review of the sixteen points fingerprint standard in England and Wales. J Forensic Identification.1996;46:49–73. Available:http://www.thefingerprintinquiryscotland.org.uk/inquiry/files/DB_0769-02.pdf.

[5] S. Gutowski. Error rates in fingerprint examination: the view in 2006. The Forensic Bulletin Autumn. 2006;2006:18–19.

[6] G. Langenburg. A Performance study of the ACE-V process: a pilot study to measure the accuracy, precision, reproducibility, and the biasability of conclusions resulting from the ACE-V process. J Forensic Identification. 2009;59(2):219–257.

[7] B.T. Ulery, R.A. Hicklin, J. Buscaglia, M.A. Roberts. Accuracy and reliability of forensic latent fingerprint decisions. Proc Natl Acad Sci USA. 2011;108(19):7733–7738. Available: http://www.pnas.org/content/108/19/7733.full.pdf. (PMC free article) (PubMed)

[8] S. M. Stigler. In The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Belknap of Harvard UP, 1986.

[9] A.B. Hepler, C.P. Saunders, L.J. Davis, J. Buscaglia. Score-based likelihood ratios for handwriting evidence. Forensic Science International, 219: 129140. doi:10.1016/j.forsciint.2011.12.009, 2012.

[10] A. Mishra. (2017). Forensic Graphology : Assessment of Personality, 4(1), 1–4. https://doi.org/10.15406/frcij.2017.04.00097

[11] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, J. R. Stat. Soc. Ser. C: Appl. Stat. 53 (2004) 109–122.

[12] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship, J. R. Stat. Soc. Ser. C: Appl. Stat. 57 (2008) 329–341.

[13] R. Marquis, S. Bozza, M. Schmittbuhl, F. Taroni, Handwriting evidence evaluation based on the shape of characters: application of multivariate likelihood ratios, J. Forensic Sci. 56 (2011) S238–S242.

[14] S.N. Srihari, S.H. Cha, H. Arora, S. Lee, Individuality of handwriting, J. Forensic Sci. 47 (2002) 856–872.

[15] T. Ali, L. J. Spreeuwers, and R. N. J. Veldhuis. A review of calibration methods for biometric systems in forensic applications. In 33rd WIC Symposium on Information Theory in the Benelux, Boekelo, Netherlands, pages 126–133, Enschede, May 2012. WIC.

[16] N. Suki, N. Poh, F. M. Senan, N. A. Zamani, M. Z. A. Darus, On the reproducibility and repeatability of likelihood ratio in forensics: A case study using face biometrics, 2016.

[17] B.T. Ulery, R.A. Hicklin, J. Buscaglia, & M.A. Roberts. (2012). Repeatability and reproducibility of decisions by latent fingerprint examiners. PLoS ONE, 7(3). Article ID e32800. http://dx.doi.org/10.1371/journal.pone.0032800

[18] F. Taroni, S. Bozza, A. Biedermann, & C. Aitken. Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. Law, Probability and Risk, 2015.

[19] J. M. Curran. Statistics in forensic science. Wiley Interdisciplinary Reviews: Computational Statistics, 1(2):141–156, 2009.

[20] J. Buckleton, C. Triggs, and C. Champod. An extended likelihood ratio framework for interpreting evidence. Science & Justice, 46(2):69 – 78, 2006.

[21] C. Champod and D. Meuwly. The inference of identity in forensic

speaker recognition. Speech Communication, 31(23):193 –203, 2000.

[22] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, and J. Ortega-Garcia. Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. Forensic Science International, 155(23):126 – 140, 2005.

[23] N. Brummer, L. Burget, J. Cernocky, O. Glembek, F. Grezl, M. Karafiat, D. van Leeuwen, P. Matejka, P. Schwarz, and A. Strasheim. Fusion of heterogeneous speaker recognition systems in the stbu submission for the nist speaker recognition evaluation 2006. Audio, Speech, and Language Processing, IEEE Transactions on, 15(7):2072–2084, Sept 2007.

[24] S. Pigeon, P. Druyts, and P. Verlinde. Applying logistic regression to the fusion of the nist'99 1-speaker submissions. Digital Signal Processing, 10(13):237 – 248, 2000.

[25] J. Gonzalez-Rodriguez, P. Rose, D. Ramos, D. Toledano, and J. Ortega-Garcia. Emulating dna: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. Audio, Speech, and Language Processing, IEEE Transactions on, 15(7):2104–2115, Sept 2007.

[26] N. Brmmer and J. du Preez. Application-independent evaluation of speaker detection. Computer Speech & Language, 20(23):230– 275, 2006. Odyssey 2004: The speaker and Language Recognition Workshop Odyssey-04 Odyssey 2004: The speaker and Language Recognition Workshop.

[27] A.O. Abiodun, A.B. Adeyemo. (2019) An Exhaustive Mapping Model for Modified Likelihood Ratio for Handwriting Recognition in Forensic Science. J Forensic Sci Criminol 7(3): 301 ISSN: 2348-9804