Stability of Individuals in a Fingerprint System across Force Levels

An Introduction to the Stability Score Index

Kevin O'Connor, Stephen Elliott
International Center for Biometric Research
Department of Technology, Leadership and Innovation
Purdue University
West Lafayette, IN, USA

Mathias Sutton, Michael Dyrenfurth
Department of Technology, Leadership and Innovation
Purdue University
West Lafayette, IN, USA

***Abstract*—This research studied the question: "Are all individual's performance stable in a fingerprint recognition system?" The fingerprints of 154 individuals, provided at different force levels, were examined using the biometric menagerie tool, first coined by Doddington et al. in 1998. The Biometric Menagerie illustrates how each person in a given dataset performs in a biometric system, by using their genuine and impostor scores, and providing them a classification based upon those scores. This research examined the biometric menagerie classifications across different force levels in a fingerprint recognition study to uncover if individuals performed the same over five force levels. The study concluded that they did not, and a new metric has been created to quantify this phenomenon. As a result of this discovery, the new metric, Stability Score Index is described to showcase the movement of individuals in the menagerie.**

Keywords—***biometrics; fingerprint; biometric menagerie; stability; instability***

I. Introduction

Integrators and algorithm developers use multiple performance analysis tools to configure biometric systems. The matching scores of individuals is important when examining what is causing errors and impacting the performance of the biometric system.

Variability in the matching scores of users is critical to all researchers of examining matching algorithms who want to choose algorithms that yield distributions with short tails [1]. Currently, commonly used biometric performance measurements are not capable of illustrating the variability amongst algorithms or different biometric systems at the individual subject level. The various methods that have been developed to classify performance, based on matching scores, all have weaknesses.

II. Literature Review

People are identified by what they have and how they act. What they have consists of traits that they are born with and should always possess. These traits are biological characteristics. Behavioral characteristics are traits that people learn or develop over time, such as writing signatures. Either of these types of characteristics is considered a biometric property, but a biometric must contain the following features: universality, uniqueness, permanence, collectability, performance, acceptability, and circumvention [2]. All of these characteristics are important when examining a biometric modality system.

A. Displaying Biometric Performance

In the biometric literature [3–5], there are four primary methods of displaying and discussing performance. These methods center around tradeoff's – whether they are between false match rates (FMR) and false non-match rates (FNMR) or false accept rates (FAR) and false reject rates (FRR). The results are then graphically displayed in score histograms, ROC curves, and DET curves. Score histograms represent the frequency in which the genuine and impostor scores occur. Receiver operating characteristic (ROC) curves graphically show the tradeoff between the verification rate and the false match rate (FMR). The detection error trade-off curves are similar to ROC curves. Instead of the verification rate represented on the y-axis, the DET curves use the false non-match rate. The majority of biometric literature discuss evaluation results in terms of these DET and ROC curves. However, an important observation relating to ROC curves is that of the area under the curve (AUC). The AUC provides an indication of performance across all values of specificity. Thus, if we compare more than one ROC curve with the same AUC, the ROC curves may not be identical. This is a weakness of the approach; the curve is simply a snapshot of the data treated as a whole.

B. Biometric Menagerie

The ROC and DET curves are graphical representations of performance illustrated by tradeoffs. However, these curves do not show information about an individual's performance. This weakness is significant because the curves do not provide the whole story, and the data may be misinterpreted. An answer to this weakness was the development of the biometric menagerie. This methodology provided additional clarity by classifying individuals by their performance. This is important because some people may contribute more error to the system than others. The zoo menagerie was popularized by [6] who coined the following animals: sheep, wolves, lambs and goats. Others have suggested alternatives, e.g., [7] and [8]. As part of their research into this problem, [7] posed the following research questions: What is the relationship between a user's genuine and impostor match scores? Does this relationship

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

exist across different biometric modalities such as the fingerprint and iris? Is there a possibility of exposing weaknesses in the biometric algorithms (i.e., comparing one algorithm with another) to see their different match rates?

Dunstone and Yager defined a new zoo menagerie, with four new animal, doves, chameleons, worms and phantoms. Doves, the best performing individuals, are in both the top 25% of the genuine distribution and the bottom 25% of the impostors. Chameleons are be in the top 25% of the genuine distribution and the top 25% of the impostor distribution. Chameleons look similar to others classifications in the dataset, as well as to themselves. Phantoms are in the bottom 25% of both the genuine and impostor distributions. These individuals are not easy to match against anyone in the dataset, including themselves. Worms, which are the worst performing. This classification occupies the bottom 25% of the genuine matches, and the top 25% of the impostor matches, indicating they do not look similar to themselves but look similar to others. However, in the zoo plots, there is a fifth category that is absent from the discussion – that of the "normal" individual. The "normal" classification describes those individuals who lie in the second and/or third quartile of the average genuine and impostor score distribution in the dataset. An example of the Biometric Menagerie is shown in Fig. 2.

The remainder of the paper is as follows. A description of the experimental methodology is presented next, followed by the results. These results show that individuals move across animal classifications, as well as within. However, just classifying them as such does not put a quantifiable number to the movement, and thus the remaining sections will discuss the development of the stability score index.

III. Methodology

Data were collected from a fingerprint recognition study that examined the impact of different force levels on the performance of the system. Five different Newton (N) force levels were used (5N, 7N, 9N, 11N, and 13N). The fingerprints of the subjects were captured on a commercially available 10-print capture device. The research question in that study was to determine the optimal force level for automated capture of high fidelity fingerprints. The data collection process required the subject to first place their four fingers on the platen from their right hand, then place their right thumb, and then the left hand and left thumb. The placement of the four fingers and thumbs was captured using the following procedure: default auto capture mode and auto-capture at 5N, 7N, 9N, 11N, and 13N.

For this analysis, individuals were identified by a Subject Identification Number (SID). Each SID was required to have 150 images (10 fingers, three placements, five force levels). If not the subject was excluded from analysis. After discarding subject data that contained missing prints or incorrect hand placement, the pool of individuals was 154. The data examined only the performance of the right index (RI) finger due to time constraints.

A. *Calculation Methodology*

The primary focus of the research was to examine the stability of individual's recognition performance with respect to force. Initially, genuine and impostor scores were calculated to understand the performance of the individual. After processing the genuine and impostor scores, the genuine and impostor distributions were averaged for each individual. The results were then plotted as *x* and *y* coordinates on biometric zoo menagerie plots.

IV. Results

The scores for all zoo plots were normalized across all five force levels. The following parameters are the standardized maximum and minimum coordinates for the zoo plots:

- Minimum Genuine (*x*-axis): 44
- Maximum Genuine (*x*-axis): 1950
- Minimum Impostor (*y*-axis): 2.4
- Maximum Impostor (*y*-axis): 10.3

Fig. 3 to Fig. 7 illustrate the zoo plots. Each graph illustrates the relative positioning of all 154 individuals with their average genuine and impostor scores. Examining each plot in sequence, readers can see the movement of the individual. This change discovers an intriguing phenomenon – that individuals move from classification to classification. Table 1 below illustrates this.

From Table 1, the total number of chameleons (those that are in the top 25% of both distributions) changes from force level to force level. This is the same for the other zoo animals. It should be noted that for 5N Normal, the 119 individuals may not be the same 119 individuals in the 11N group. Thus, this result prompted the authors in the development of a metric to describe this phenomenon, namely the Stability Score Index.

A. *Stability of a Zoo Menagerie Animal*

No individuals were able to obtain the same genuine and impostor scores across force levels, but some showed significantly smaller movements in the zoo plots. Only one individual was classified as the same zoo animal in all five force levels. The other 153 individuals were classified a different zoo animal on at least one of the force levels. The different classifications do not necessarily mean the individual is unstable. Furthermore, if an individual is classified the same on all force levels this then doesn't mean that the individual was "stable" either. For instance, individual 034 was classified as a chameleon on all force levels. Their movement was still variable, even within the chameleon classification.

TABLE I. Animal Classification Changes over Force Levels

Table Head	**5 N Count**	**7 N Count**	**9 N Count**	**11 N Count**	**13 N Count**
Chameleons	11	16	22	15	16
Doves	5	5	9	6	6
Normal	119	114	102	119	117
Phantoms	12	16	16	13	11
Worms	7	3	5	1	4
Total	154	154	154	154	154

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

B. Stability of the Normal Classification

In previous research, researchers have tended to ignore that "normal" classification. For example, [3] defined the new animal classifications but ignored the "normal" classification, referred to in their papers as the "none" classification. However, in this study, the majority of individuals are within the normal group, which creates the opportunity for the individual to move significantly without changing classification. Thus, it is an important classification to examine.

The "normal" classification of individuals lies in the second and/or third quartile of at least one of the genuine or impostor scores in the dataset. If an individual performs consistently in this "normal" classification, this should not be ignored. The "normal" shows that the current animal classification is not adequate because the "normal" classification comprises of the majority of the zoo plot. Our analysis indicates that individuals exhibit some instability within this classification. Individual 135 was examined for this topic due to their performance of instability and stability in the "normal" classification. From force levels 5N to 7N their average genuine score changes from 485.6667 to 1155, while maintaining the same classification. From 9N to 13N, their match scores become stable.

C. Borderline Individuals

When examining the zoo plots, it is evident that there are individuals that are borderline cases. This can be seen as a weakness of zoo plots. For example, Fig. 1 illustrates two individuals that have similar genuine and impostor averages but classified differently.

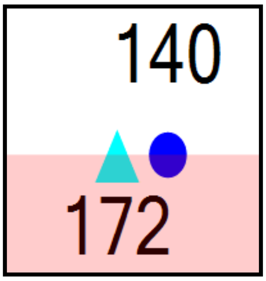

Fig. 1. Borderline case at 13 N for individuals 172 and 140.

Individuals 172 and 140 are classified as a chameleon, and as a "normal", respectively. This is because they have slightly different impostor scores: 9.0675 and 9.0661, respectively. If these individuals were to take assume each other's impostor scores at a different force level, they would reverse their classifications. Examples such as this prompted our new approach to calculating individuals' movement independent of their animal classification.

D. Stability across Zoo Menagerie Animals

There are two cases of stability across zoo menagerie animal classifications. The first case is having an individual move in smaller segments and still be classified different. This is similar to borderline cases where subjects can be stable but are deviating close to a border where they are classified different across force levels. For example, individual 178 had relatively similar genuine and impostor scores across the 7 N and 9 N zoo plots. The weakness by just examining the animal classification is the individual would appear to have an unstable performance, due to being classified differently.

The biggest instability cases also involve a change in zoo animal classification. Individual 117 was classified as a normal on the 5N, 7N, and 13N zoo plots. From 9N to 11N, individual changes from a dove to a phantom. This movement and change from classification showed a large change in the average genuine score from 1526.33 to 471.0.

V. Stability Score Index

The motivation behind the stability score index was to address the movement from one classification to another, or within a classification (something that cannot be seen from the aggregate zoo analysis). Conceptually, the stability score index examines the movement of the subject across two different datasets – for example from 5N to 7N. If the subject does not move "much", we deem them "stable". If they move "a lot", then we deem them unstable. To quantify, a stability score ranges from 0 to 1. In this case, 0 is stable (i.e., the individual does not move at all). One is unstable, where the subject moves the maximum difference. The stability score index (S.S.I), shown in Eq. (1), is used to calculate the stability for any individual (i) of interest from one force level to the next.

$$S.S.I_i = \frac{\sqrt{\left(x_{i_2} - x_{i_1}\right)^2 + \left(y_{i_2} - y_{i_1}\right)^2}}{\sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2}} \quad (1)$$

The approach is described as the following: x_1 and x_2 represent the genuine scores for the two force levels examined. y_1 and y_2 represent the individual's impostor scores from each force level. $x_{\max}$ and $x_{\min}$ represent the maximum genuine score and minimum possible score that occurred across all force levels. $y_{\max}$ and $y_{\min}$ represent the maximum impostor score and minimum possible score at all force levels. The numerator value represents the individual's movement from the two force levels and the denominator the maximum possible change in all force levels.The force variable can be substituted for other variables such as time, multiple sensors, or multiple modalities.

A. Stability Score Index for Subject 34

Individual 34 was discussed in a previous section regarding stability within the same classification on all five force levels. Instability can occur within a classification at different force levels. An individual is capable of moving ¼ of the maximum possible movement and remain in the same classification. In the examined data, the maximum movement was not observed, but an instance of smaller movements showed that the possibility exists. Table 2 is a list of all the examined stability score index results for individual 34.

B. Stability Score Index for Subject 135

Individual 135 was also examined previously for instability within the normal classification. The stability score

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

and related coordinates for the 5 N and 7 N levels were as follows: as follows: the 5 N genuine score is X_1 (485.6666), the 5 N impostor score is Y_1 (7.0901), the 7 N genuine score is X_2 (1155), and the 7 N impostor score is Y_2 (8.6005). The value thus obtained is 669.335, which is divided by the maximum movement of 1906.0164 to give a stability score index of 0.3512.

Table 3 is a list of all the examined stability score index results.

C. Stability Score Index for Individual 117

The most drastic case of instability involves a change in animal classification. Individual 117's classification changes from a dove to a phantom at different force levels. For individual 117, both the zoo plots and the stability score reflect a high level of instability. On 9 N, individual 117 is classified as a dove and a phantom at 11 N. The stability score should reflect the great movement at different force levels. By using the coordinates to calculate the stability score index, a value of 0.5537 is obtained.

Table 4 is a list of all the examined stability score index results.

D. Stability Score Index for Individual 178

Individual 178 was examined as a stable performance across force level but was assigned different classifications. This weakness of the zoo plot is compensated for with the stability score index. Table 5 shows the small deviation from the 7 N results to the 9 N results. Regardless the classification for individual 178 in the zoo plots, the stability score remains the same, close to zero, indicating stability. Inserting the coordinates into the formula, a stability score of 0.0308 was obtained.

E. Stability Score Index Conclusion

The stability score index does not use the classification methods that have been proposed in the literature, but focuses on individual performance from a discrete perspective. Fig. 8 graphically represents all stability scores from each individual across the five force levels in the following manner: 5 N to 7 N, 7 N to 9 N, 9 N to 11 N, and 11 N to 13 N. There can be numerous additional combinations, but this research is limited to the described relationships.

VI. SUMMARY

The results of this research show the presence of instability in individuals in fingerprint recognition for the right index finger. Three different cases have provided evidence that individuals are unstable. Only four individuals were highlighted due to space constraints, but many others in this particular dataset showcase inconsistent stability throughout the five different force levels.

TABLE II. INDIVIDUAL 034 SSI TABLE

SID	5-7N	7-9N	9-11N	11-13N
34	0.0264	0.1296	0.0175	0.0462

TABLE III. INDIVIDUAL 135 SSI TABLE

SID	5-7N	7-9N	9-11N	11-13N
135	0.3512	0.0065	0.0633	0.0275

TABLE IV. INDIVIDUAL 117 SSI TABLE

SID	5-7N	7-9N	9-11N	11-13N
117	0.2006	0.1995	0.5537	0.1658

TABLE V. INDIVIDUAL 178 SSI TABLE

SID	5-7N	7-9N	9-11N	11-13N
178	0.0967	0.0308	0.2723	0.0894

This instability can result from many underlying variables such as the quality of images because of the force, subject familiarity with the fingerprint sensor, or randomization of the levels at which the individual was tested. More research on this topic should be conducted, to see whether this phenomenon described in the paper is evident in other modalities.

REFERENCES

[1] M. Shuckers, "Additional topics and discussion," in *Computational Methods in Biometric Authentication*, M. Shuckers, Eds. New York: Springer, 2010, pp. 299–300.

[2] A. Jain, R. Bolle, and S. Pankanti, "Introduction to biometrics," in *Biometrics: Personal Identification in Networked Society*, Springer, 2002, pp. 141.

[3] T. Dunstone, and N. Yager, *Biometric system and data analysis*. New York, NY: Springer Science+Business Media, 2009.

[4] ISO / IEC JTC 1 SC 37, Text of FCD 19795-1, *Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, (N908). Geneva, 2005.

[5] J. Wayman, A. K. Jain, D. Maltoni, and D. Maio, "An introduction to biometric authentication systems,". in *Biometric Systems*, J. Wayman, A. Jain, D. Maltoni, and D. Maio, Eds. London: Springer, 2005, pp. 1–20.

[6] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *Proceedings of the 5th International Conference on Spoken Language Processing*, 1998, pp. 1–5.

[7] N. Yager, and T. Dunstone, "The biometric menagerie," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 220–30, 2010.

[8] E. Tabassi, "Image specific error rate: A biometric performance metric," in *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, pp. 1124–1127. doi:10.1109/ICPR.2010.281

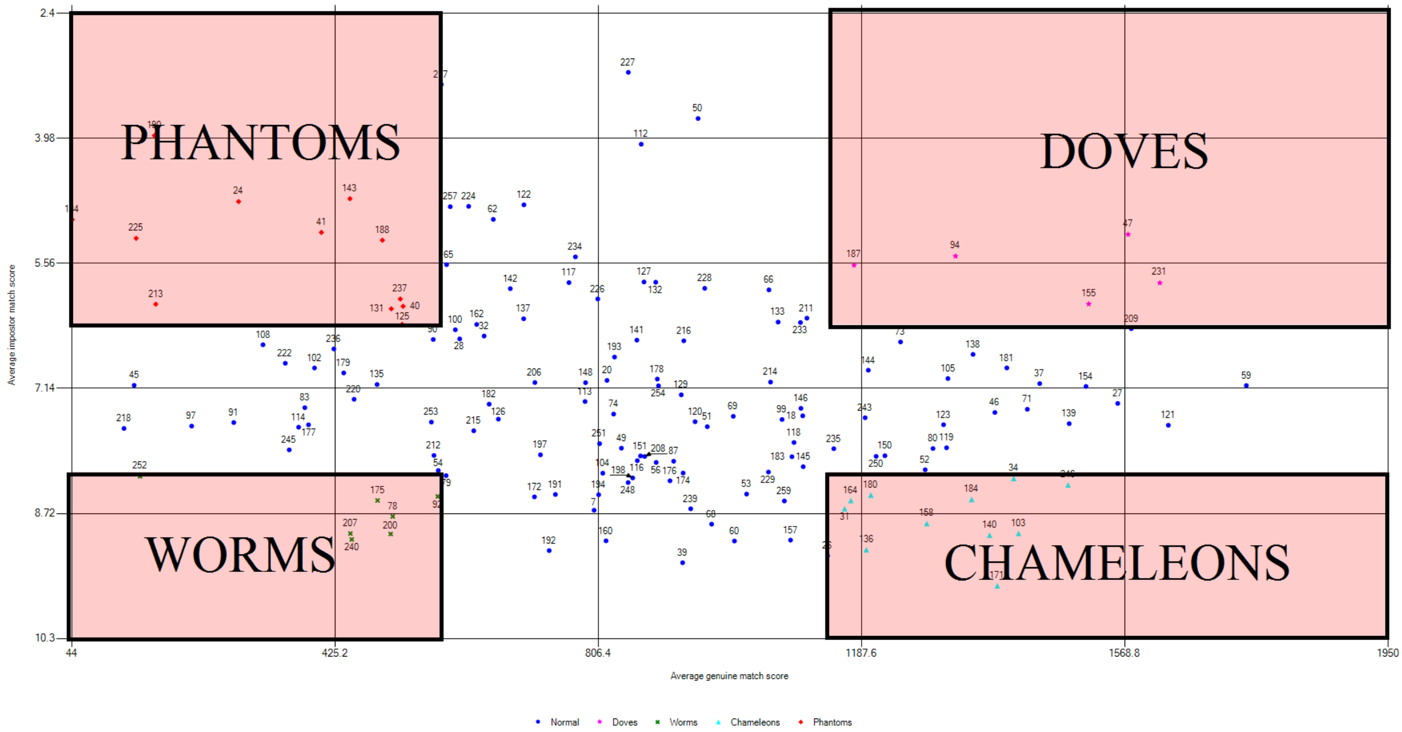

Fig. 2. Biometric Zoo Menagerie Plot Example.

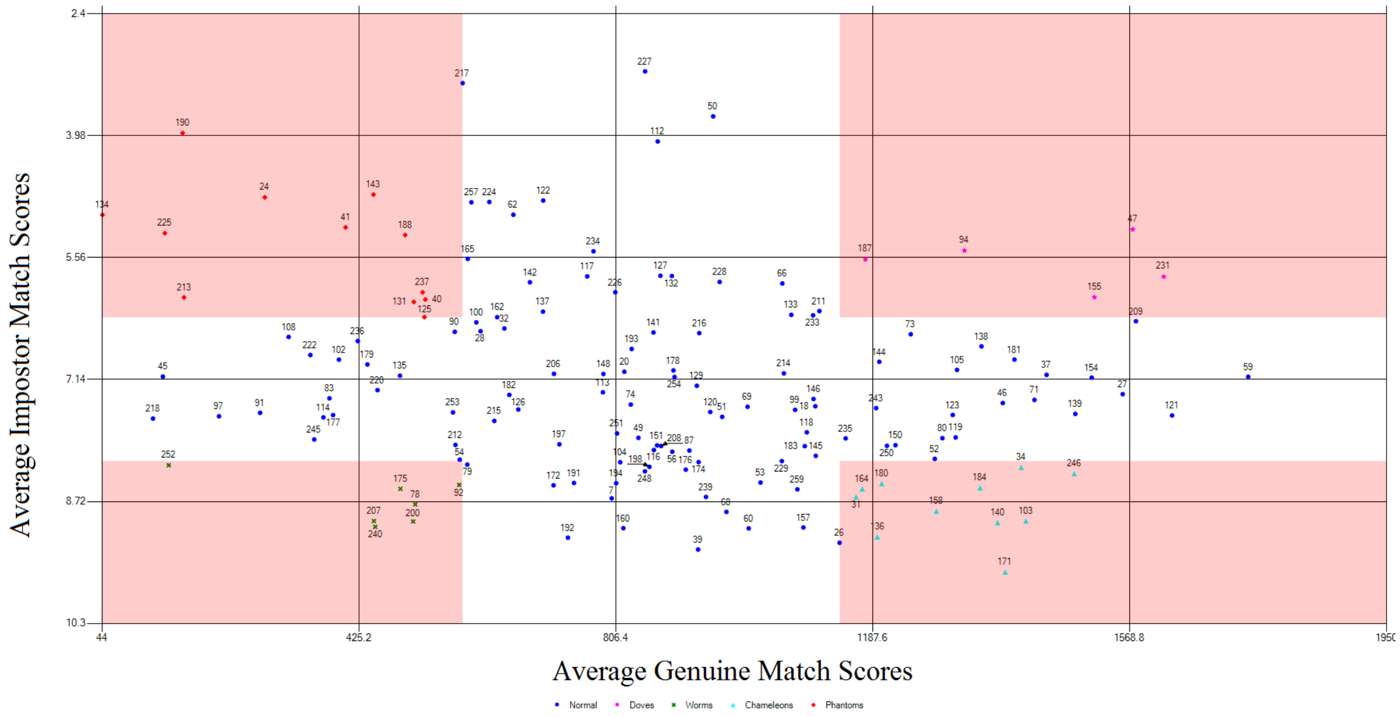

Fig. 3. 5 N Zoo Plot.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

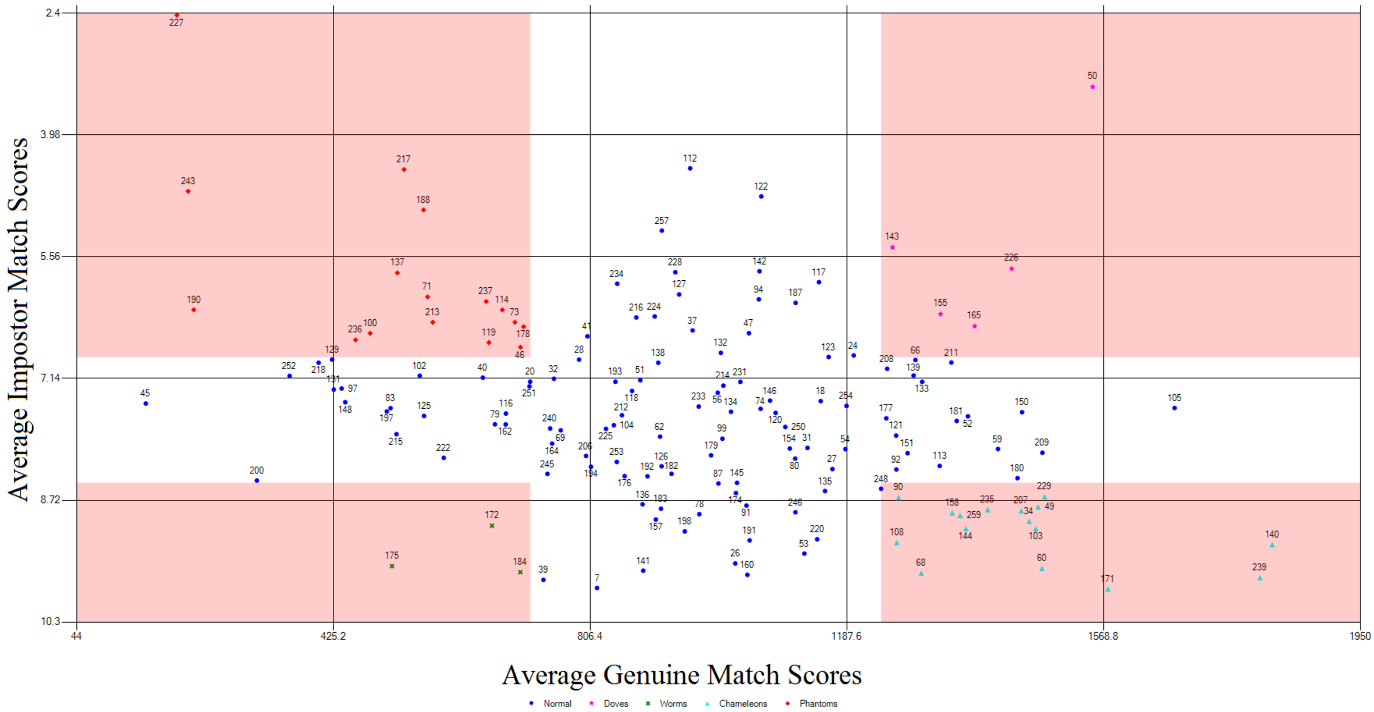

Fig. 4. 7 N Zoo Plot.

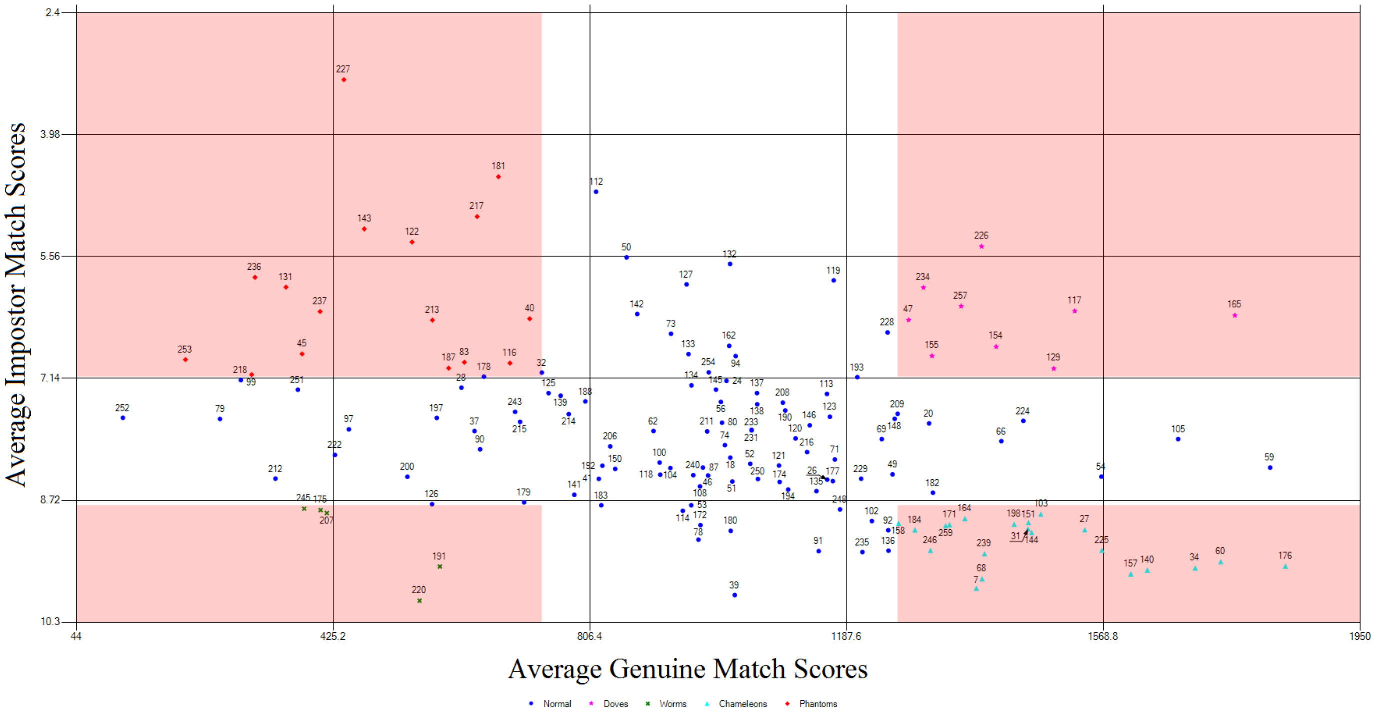

Fig. 5. 9 N Zoo Plot.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

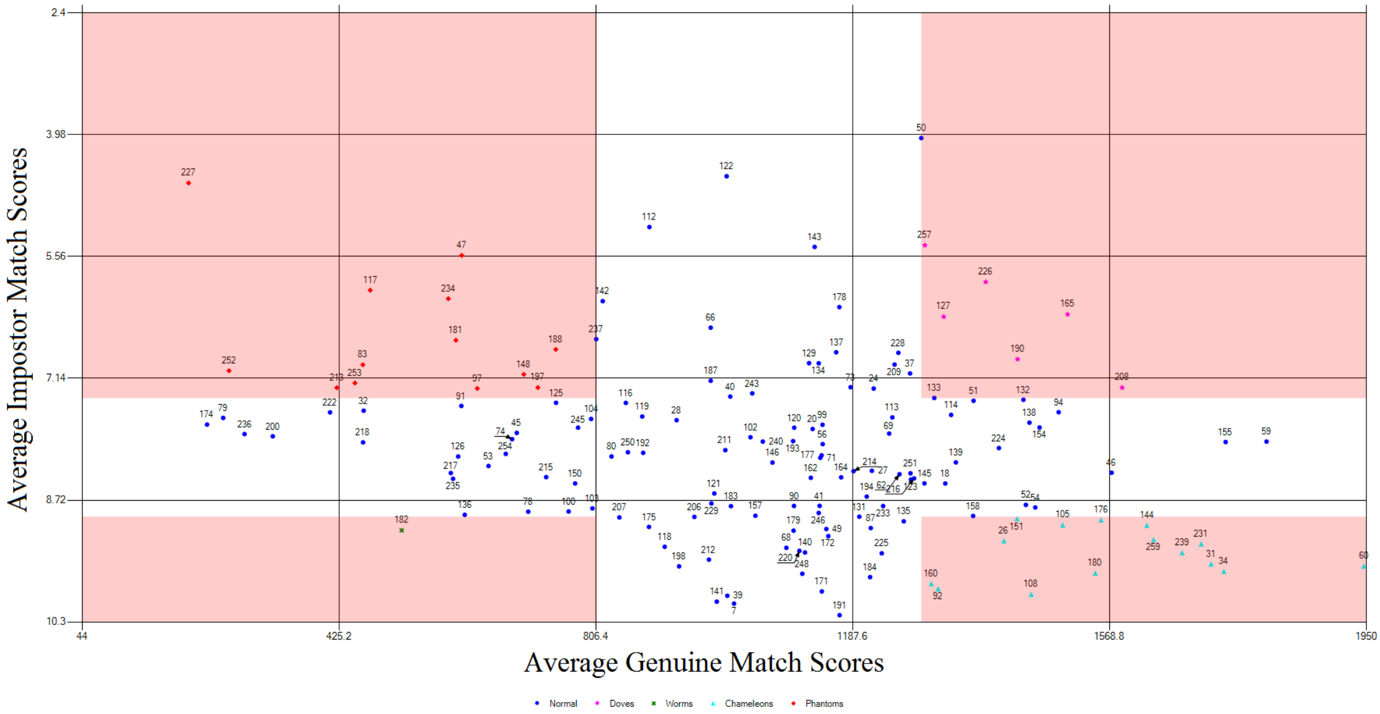

Fig. 6. 11 N Zoo Plot.

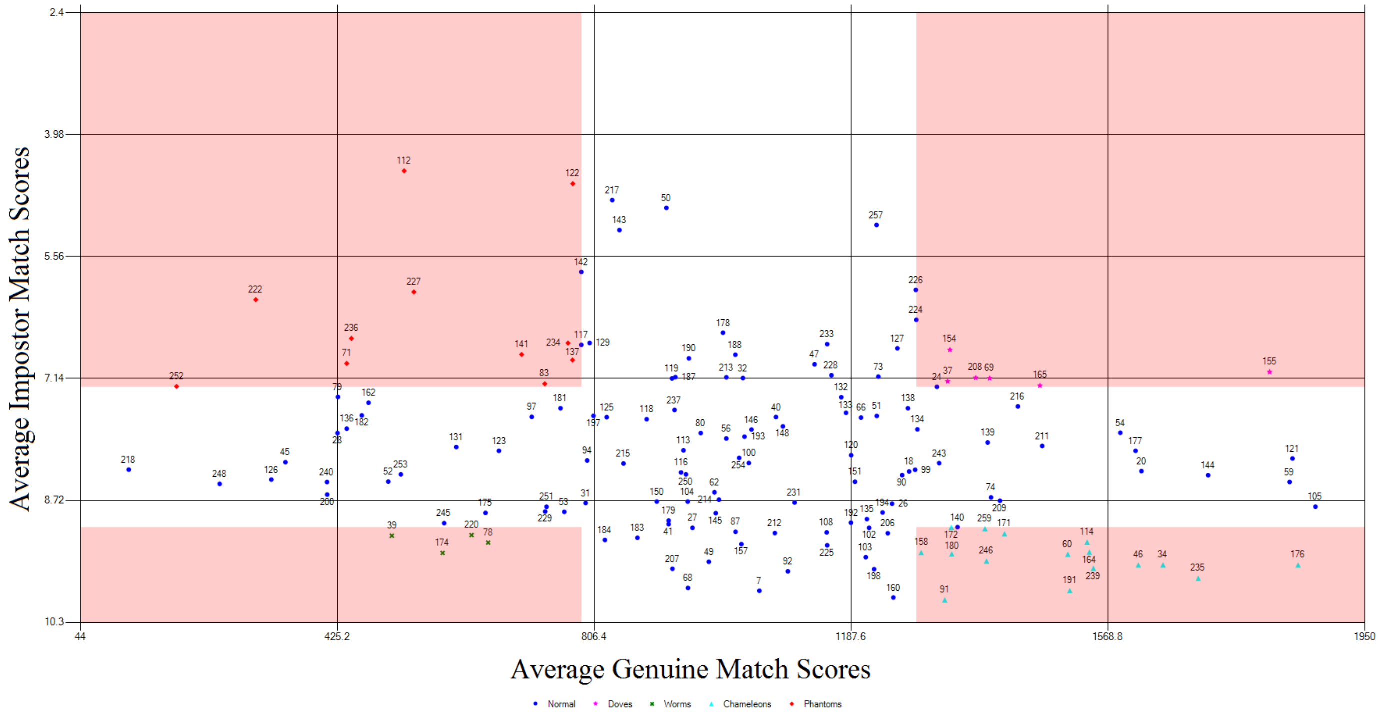

Fig. 7. 13 N Zoo Plot.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

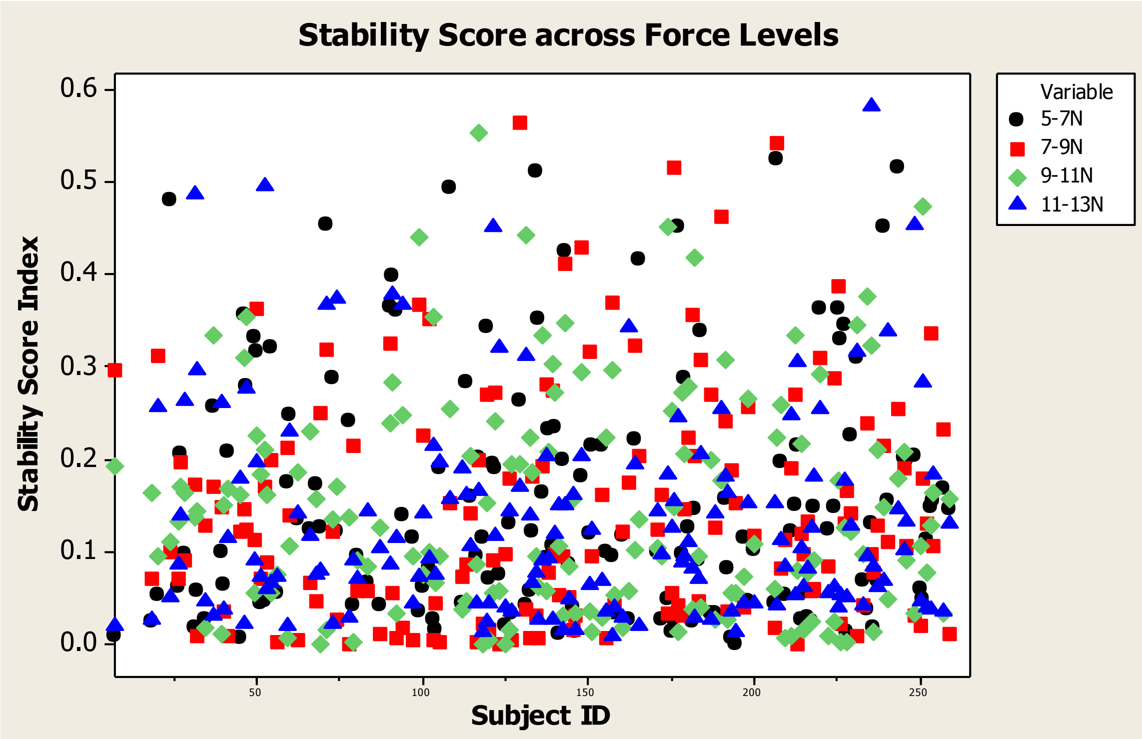

Fig. 8. Scatterplot of stability scores for each individual.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731