Investigating the use of feature selection techniques for gender prediction systems based on keystroke dynamics

NASCIMENTO, Tuany Mariah Lima Do, OLIVEIRA, Andrelyne Vitoria Monteiro De, SANTANA, Laura Emmanuella Alves dos Santos and DA COSTA ABREU, Marjory <http://orcid.org/0000-0001-7461-7570>

Investigating the use of feature selection techniques for gender prediction systems based on keystroke dynamics

Tuany M. L. Nascimento, Andrelyne V. M. Oliveira, Laura E. A. S. Santana[1], Márjory Da Costa-Abreu[2]

[1]Federal University of Rio Grande do Norte, Natal, BR
[2] Department of Computing, Sheffield Hallam University, Sheffield, U.K.

Keywords: gender, keystroke, feature selection, metaheuristics, genetic algorithms.

Abstract

Biometric-based solutions keep expanding with new modalities, techniques and systems being proposed every so often. However, the first ones that were used for authentication, such as handwritten signature and keystroke dynamics, continue to be relevant in our digital world, despite their analogical origin. In special, keystroke dynamics has had an increase in popularity with the advent of social networks, making the need to continue to authenticate in desktop or game-based user verification more prevalent and this became an open door to risky situations such as paedophilia, sexual abuse, harassment among others. One of the ways to combat this type of crime is to be able to verify the legitimacy of the gender of the person the user is typing with. Despite the fact that keystroke dynamics is well accepted and reliable, this technique can have far too many attributes to be analysed which can lead to the use of redundant or irrelevant information. Therefore, propose a comparative study between two features selection approaches, hybrid (filter + wrapper) and wrapper. They will be tested by using a genetic algorithm, a particle swarm optimisation, a k-NN, a SVM, and a Naive Bayes as classifiers, as well as, the Correlation and Relief filters. From the results obtained, it can be said that the two proposed hybrid approaches reduce the number of attributes, without negatively impacting the accuracy of the classification, and being less costly than the traditional PSO.

1 Introduction

The widespread popularity of the use of biometric data has created new opportunities to guarantee security and authentication/verification reliability [13]. One of the possible applications is instant messaging such as Whatsapp [1], Telegram [2] and Signal [3]. Its main aim it to serve to unite different groups of people and also to shared information [12]. However, users may be vulnerable to risky situations, for example, pedophilia and insistence [9]. One of the ways to investigate a suspected pedophile is to confirm whether the individual is, in fact, the sex that it is said to be [9]. Therefore, a viable technique to use is the keystroke dynamics [1].

Keystroke dynamics is a behavioral biometrics technique, that analysing how the user types in a terminal and monitors the keyboard to recognize the user based on their usual rhythmic patterns [22]. Data about keystroke dynamics is often taken in through free text or fixed text, this is, texts that have periodic monitoring of the keystrokes and text predetermined for all people under observation (login and password for example), respectively [6]. This kind of information is accessible to be widely used since no extra external hardware, apart from the keyboard, only required to collect data [26].

Works Previous show that keystroke dynamics can be used technique for gender recognition [32], however, this approach can extract many features and the possibility of needing to deal with a large number of features in a classification task can have a negative impact on performance [8]. Intuitively way, the greater the number of features in a dataset, the easier it is to extract knowledge models, however, in practice, this is not be true in most cases, since there could also be redundant, irrelevant features or with presence of noise that can confuse the classifier as well as impair its accuracy [8].

With the very large number of features in a database, high computational power is required and this makes it difficult to build the model [8]. Thus, the features selection has been one of the fundamental steps of data preprocessing for other applications in tasks such as pattern recognition, data mining and machine learning, among others, with its main objective being to select a subset of relevant features among all the features available to the task [28].

Feature selection is used with two approaches: wrapper or filter [29]. In the work of [32] involve gender recognition for problems using keystroke dynamics, feature selection with the filter approach, where, the authors used an evaluation metric the information gain.

Studies with the use of the wrapper and filter approaches for features selection-based on keystroke dynamics but for individuals identification, as [21] where the authors used as metaheuristics, genetic algorithms and as classification algorithms k-nearest neighbour (kNN), support vector machine (SVM), LibSVM and multilayer perceptron (MLP) and Wisard, to confirm their results, they used the clustering techniques, k-means. In the work of [27], the authors proposed a comparative analysis between genetic algorithms and particle swarm optimiza-

[1]www.whatsapp.com/
[2]https://telegram.org/
[3]https://signal.org/

tion to feature selection in keystroke dynamics databases for individual identification, the classifier used by them was the random forest that presented a result 92.58% and it was observed that the database reduced from 49 features to a range of 17 to 26 features.

The work of [17] proposes the combination of five filters, which were: Relief, F -score, MI, mRMR and Rough Set, using the PSO metaheuristic to produce a new hybrid filter wrapper approach, for feature selection in the Arcene, Madelon, Mushroom and Spambase datasets.

Thus, this work proposes performing a comparative analysis between meta-heuristics models performing feature selection gender recognition using keystroke dynamics and a wrapper approach with either particle swarm optimisation and genetic algorithms, and two-hybrid approach applying filters to rank the features in the step the crossover of genetic algorithm and in the initial population, respectively.

The paper is organized as follows: Section 2, we will present the main aspect of keystroke dynamics and gender prediction. In Section 3, we will address concepts of genetic algorithm and its use for feature selection, as well as its proposed hybrid forms. In Section 4, we will present the methodology applied in this work with a description of the used databases, as well as the classifiers and filters used, and finally, in Section 5, we will present our conclusions on the results obtained from the experiments.

2 Keystroke dynamics used for gender prediction

Since our work aims to compare the wrapper and hybrid approaches for feature selection in keystroke dynamics datasets for gender recognition, we need to understand what has been done so far regarding keystroke dynamics and gender recognition.

Keystroke dynamic is also known as keyboard-dynamics, is considered one of the techniques of behavioural biometrics and has had the purpose of identifying users based on the typing patterns [22]. There are many features that can be obtained from this, such as: first key-press time, flight time between key presses and second key press time.

The identification of a user through keystroke dynamics it occurs by comparing an individual's current activity with other similar ones stored, in order to detect significant differences in patterns to prevent unauthorized users from having access to other people's data, using computers [9].

The methods for analysing the user's keystroke dynamics can be divided into two approaches: fixed text and free text. The fixed text authentication takes place using predefined words or texts, such as login and password, making it possible to capture specific features of a specific person. However, free text authentication does not require specific texts for the collection of user data, that is, pressure analysis is based on any text in the individual's daily life, requiring constant monitoring [16]. The data it is obtained through the keyboard, without the addition of hardware [33].

Because it is a behavioral method, keystroke dynamics can show variation in de typing users over time, in relation to physical biometrics, even without considering the physiological or psychological state of the individual. [10].

As mentioned previously, the feature selection is taken into account to be amongst the foremost important steps within the process of data mining, machine learning and pattern recognition, since it aims to reduce dimensionality by selecting a subset of features that are relevant to the building of a model [28].

Feature selection can be used two approaches: filter or wrapper. The wrapper approach [29] this approach requires high computational power and uses the accuracy of the classifier to evaluate the subset of selected features [30]. The filter approach act independently of the learning algorithm and some evaluation criteria which are applied, assigning for each feature. The evaluation criteria are defined based on the information gain, the correlation and the distance between features [4]. Besides that, the filter approach has low computational cost compared to the wrapper approach [14, 20, 3].

In works such as [14], the authors proposed a hybrid two-step approach combining wrapper and filter in order to find the best subset of features. The filter approach, the first step, was used for ranking of features and in the next step, they were ordered in a high ranking in the wrapper approach with particle swarm optimisation and genetic algorithms. The filters used by the authors were minimum redundancy and maximum relevance (mRMR) and CFS in the fitness function step of the genetic algorithm and as the classifiers, they used support vector machine *(SVM)* with 4-fold-cross-validation. The results presented by the authors showed that the hybrid approach is effective for obtaining a smaller and optimal subset of features.

The literature has already demonstrated that it is possible to predict gender from biometric data, including keystroke dynamics, as it can be seen in the work of [2] where two databases were used in their experiments. For the feature selection, the authors used the filter approach and the correlation evaluation method. They used Random Forest as classifier, and they presented good results with cross-validation in both datasets.

[5], shows in his work the combination of two biometric modalities, keystroke and handwritten, for the prediction of users' gender. The dataset used by them was attended by 100 participants, 45 men and 55 women. The multimodal dataset included 400 instances and 29 features for each keystroke dynamics task, and 49 for each calligraphy task, which is described in [19]. The feature selection used by the authors was random, where the features were chosen at random and each classifier selected its subset own. The classifiers chosen by the authors were: k -NN, MLP, SVM and Decision Tree (DT).

And finally, the work presented in [24] carried out feature selection in a keystroke dynamics database for gender recognition using the metaheuristic genetic algorithm. They also used two databases being they [10] and [31], as classifiers they used k-NN, SVM and Naive Bayes, bringing SVM with 88.10 % as the best result.

From what we saw in this section, none of the studies cited, it is possible to find a significant discussion on feature selection using genetic algorithms in keystroke dynamics databases for gender recognition.

3 Using genetic algorithm (GA) as a wrapper for feature selection in keystroke dynamics

Since our proposed new systems are related to modifications in the traditional structure of genetic algorithms, this section will give a brief introduction of GA as well as present the new steps we have tested.

Inspired by Darwin's theory of evolution [7], the GA starts with a set of possible solutions to solve a given problem called the initial population. The quality of the individuals in this population is evaluated through a function. After this evaluation, the genetic operators are repeated until the population reaches the stopping condition [29].

The next sessions present the implementation of two new hybrid approaches involving genetic algorithms, section 3.1 describes the application of the filters in the crossover step and the formation of twin children, while the section 3.2 describes the application of the filters, separately, right at the initial population.

3.1 Genetic Algorithm Crossover (GAC)

In this work, the hybrid approach proposed performs feature selection in the step crossover of the genetic algorithm, in such a way that when crossing the best attributes classified by the filters, it can result in the improvement of accuracy.

In more details, in the feature selection using our first hybrid approach, which we will now call GAC (Genetic Algorithm Crossover), the algorithm follows the entire flowchart described in Figure 1. The algorithm showed in this work is the application of two filters, Relief and Correlation, in the crossover step of the genetic algorithm, where the respective filter was used in a parent, returning half of the characteristics of this filter, considering the ranking generated by the filter. With the crossover from parents must generate two children to maintain the size of the population, this child was duplicated being able to form twins. The child generated will be the union of the selected features of the two parents.

Once this is done, the GAC follows its normal flow, upon reaching the probabilistic rate in the mutation step, can cause the twins to differentiate and adding these children to a new population.

With this, with the use of filters in the crossover step, greater accuracy in the classification is expected. The implementation of this algorithm can be found in [23].

3.2 Genetic Algorithm Initial Population (GAIP)

In our second proposed approach, the application of filters was developed in the initial population, thus, the algorithm already starts with the best attributes classified by the filters and it can result in the improvement of accuracy, as will be described.

We will now call this variation as the Genetic Algorithm Initial Population (GAIP). In it, the algorithm follows the entire flowchart previously described in Figure 2. However, this approach presents the proposal of applying the Correlation and Relief filters separately in the initial population of the genetic algorithm, thus, each individual that composes the initial population already starts with its rankings.

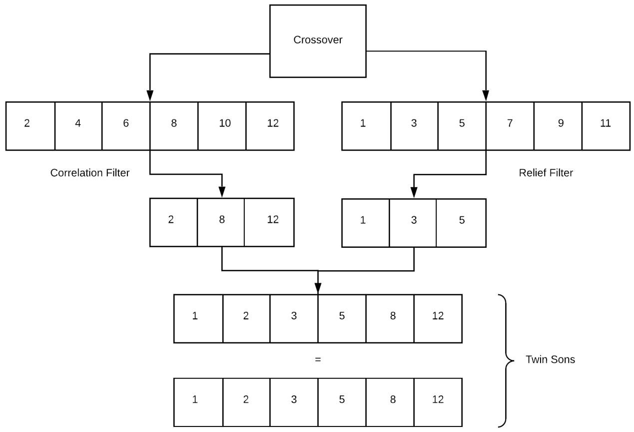

Figure 1. Hybrid Approach - GAC

Despite the existence of several works in the area, no work was found that did feature selection using genetic algorithm and its hybrid form in keystroke dynamics databases for gender recognition. Our idea was to limit the quantity of generations by giving a meta-heuristic extra information that could help focus the search area of the problem.

In the next section we will present our methodology and how we validated our proposed techniques for gender prediction using keystroke dynamics.

4 Methodology

As mentioned previously, our goal is to compare the wrapper and hybrid approaches for feature selection in keystroke dynamics datasets for gender recognition. Thus, in this session we will present the databases, as well as the classifiers and filters used in our experiments. For the development of the present work, were used two databases which are described in the Table 1.

	Volunteers	Men	Women	Features	Instances
Hand-based-database [31]	77	20	50	42	231
Giot [10]	133	98	35	60	7555

Table 1. Database Details

The databases used in this work were used in the work of [24, 22, 9] which is publicly available by contacting the authors.

To perform a comparative analysis between the feature selection techniques, we used three classifiers, that can be found in Weka toolbox: Support Vector Machine (SVM) [8], Naive Bayes [8] and k-Nearest Neighbor (KNN) [8]. The classifiers were chosen because they belong to different learning paradigms, and could bring different aspects to the comparative analysis to be performed, as well as being frequently used

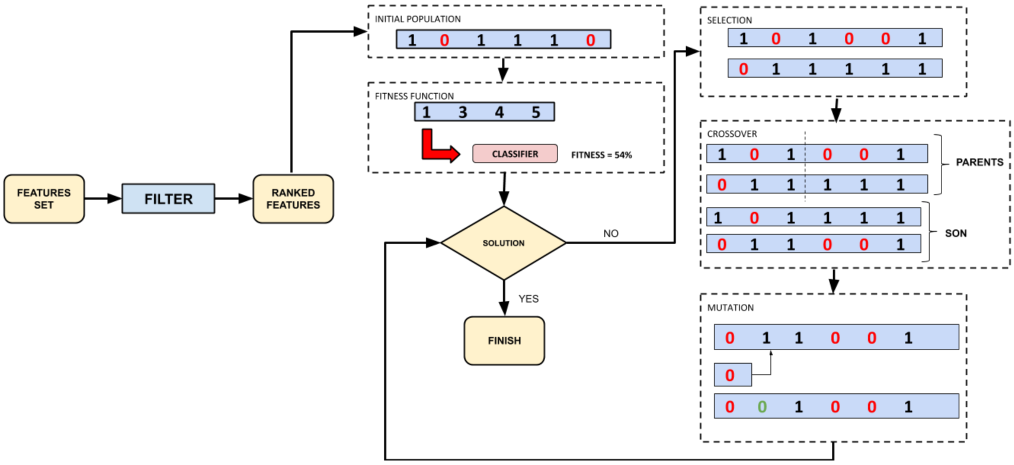

Figure 2. Hybrid Approach - GAIP

in the literature.

For comparison with our proposed GAC and GAIP, we have decided to perform experiments with Particle Swarm Optimisation (PSO). We have chosen to use the version proposed in the article by [15]. This implementation is commonly used to perform feature selection in a database using the wrapper approach [29]. For this work, we have used the discrete PSO [25, 18].

The filters used in this work are in the Weka toolbox [11]. These filters have been applied in the crossover step and the initial population, making two hybrid feature selection.

With the intention of analysing the impact of feature selection on the performance of each classifier when applied to keystroke dynamics databases for gender recognition, the different filters were used based on [22]. The filters used were the **Correlation** and the **Relief** [11].

These two filters are mostly used for feature selection and that is why we have chosen to analyse their performance in this work [14]. The next section will present our main results and the important conclusions and contributions from our proposed two new techniques.

5 Results and Conclusion

Considering that our aim is to analyse the feature selection in keystroke dynamics database for gender recognition, our experiments were performed using two databases, Hand-based-database with 43 features and 231 instances, and Giot database with 60 features and 7555 instances.

Table 2 show the accuracy of the classification with all features for the dataset divided into 66% of the training instances and 33% for the test, for the k-NN, SVM and Naive Bayes classifiers. For the confirmation of our results, we used the t-test, analysing the behavior of each classifier, separately, when used with the wrapper and hybrid (GA, GA [24] and PSO) approaches, as well as no selection to confirm if the presented result was statistically significant. The difference is significant if the test result is less than 0.05 (acceptable error probability 5%). In this case, we reject the null hypothesis, that would indicate that the difference is only apparent.

We have divided, for simplicity, our results presentation into Sections 5.1, 5.2 so we can see the changes' impact on each different dataset.

5.1 Giot database

In Table 2 we show the accuracy of the classifiers using the full base with 60 attributes is between 69.22% and 86.09%, with the KNN being the classifier that presents the best result. However, when the database is reduced by the proposed approaches, there is a noticeable increase in the accuracy of the classifiers, that is, the results improved after the feature selection. In the proposed approaches for feature selection, wrapper and hybrid, the SVM classifier presents better performance, in GA with 90.77%, GA [24] with 90.05%, PSO with 90.87%, in GAC with 90, 11% and in GAIP 90.67% and 90.74% using the correlation and relief filters, respectively.

As can be seen that in the complete database (no selection), there are irrelevant and redundant attributes that influence the performance of the classifiers and that with the feature selection more significant results are achieved. This is noticeable in the performance of the Naive Bayes classifier since, for this estimator, the correlated attributes affect negatively since it acts independently. As it can also be seen for the SVM classifier, which with the complete database has an accuracy of 69.22%, and after the feature selection its accuracy is over 90.87%, which may have occurred for the accuracy to be so low, using the complete database is overlapping classes due to the imbalance of the database.

For the **GAIP** approach, statistical tests were performed making comparisons between GA [24], GA, PSO, GAC, and no selection. Thus, it can be observed that when analysing the GAIP, there is no statistical difference between the two filters analysed, correlation and Relief. Analysing this same approach concerning GA, it can be seen that for the k-NN with 88.89% accuracy, GA obtained better performance in contrast to SVM and Naive Bayes, which had its best performance in GAIP, in both filters. New experiments were carried out, with the wrap-

	NO SELECTION		
DATABASE	KNN	SVM	NAIVE BAYES
GIOT	86.09 ± 0.54 (60)	69.22 ± 0.72	74.39 ± 0.61
HAND	60.12 ± 4.00 (43)	74.70 ± 2.91 (43)	56.14 ± 4.00 (43)
	GA		
GIOT	88.44 ± 0.25	90.77 ± 0.19	79.09 ± 0.35
HAND	87.05 ± 2.95	84.47 ± 1.95	75.57 ± 3.58
	PSO		
GIOT	88.54 ± 0.22	90.87 ± 0.13	79.00 ± 0.37
HAND	88.27 ± 1.87	84.73 ± 1.46	76.29 ± 3.36
	GAC		
GIOT	86.15 ± 0.62	90.11 ± 0.10	76.53 ± 0.41
HAND	83.08 ± 1.06	83.71 ± 0.43	81.26 ± 0.60
	GAIP - Correlation Filter		
GIOT	88.37 ± 0.24	90.67 ± 0.17	78.33 ± 0.26
HAND	88.85 ± 2.07	84.68 ± 1.77	75.50 ± 2.44
	GAIP - Relief Filter		
GIOT	88.38 ± 0.22	90.74 ± 0.18	78.29 ± 0.34
HAND	86.16 ± 1.40	82.87 ± 1.52	73.74 ± 2.07

Table 2. Accuracy (Acc) rates in percentage for the Giot and the Hand-based databases

per approach using GA, with different configurations for k-NN and SVM, with K = 1 and C = 10, as shown in Table 2, with this, it can be observed that there was no statistical difference between the two classifiers. We carried out the experiments with the PSO, which showed better results with 88.54%, 79%, and 90.87%, for the k-NN, SVM and Naive Bayes classifiers.

What can also be observed GAIP, ranking the attributes at 50%, the accuracy value is statistically similar, when compared to the complete database, what may have occurred is the small number of attributes or the presence of noisy attributes.

The **GAC** approach is compared to GAIP, PSO, GA [24] and GA with new configurations, obtaining poor performance in all scenarios for all classifiers.

5.2 Hand-based-database

In Table 2 we show the accuracy of the classifiers using the complete database with 43 attributes is between 55.54% and 73.55%, the SVM showing better results with 73.55%, the hypothesis for this is that for this database, there are not as many overlapping classes as occurred in the database [10]. However, when we perform the feature selection with the proposed approaches, on the [31] database, a significant increase in the accuracy of the classifiers is noticeable. In the proposed approaches to feature selection, wrapper, and hybrid, the classifier that presented the best performance was k-NN, in GA with 87.05%, GA [24] with 87.85%, PSO with 88.27%, GAC 88, 85% and 86.16% using the correlation and relief filters, respectively. After the feature selection, the three classifiers showed a significant increase in their accuracy, as it is possible to analyse that the complete database presents redundant and irrelevant attributes that end up affecting the performance of the classifiers negatively, this is noticeable with the increase in accuracy in the k-NN and Naive, that both are affected by the presence of irrelevant and redundant attributes.

For our proposed **GAIP** appproach,according to the statistical tests performed, it can be observed that this approach using the correlation filter obtained better performance when compared to the Relief filter. For GA [24], it brought the best result using the Relief filter, for all classifiers, however when compared to the correlation filter, only the k-NN best result. As previously stated, new experiments were carried out, with the wrapper approach using GA, with different configurations for k-NN and SVM, with K = 1 and C = 10, with this it can be seen that with the new configurations, k-NN did better using the correlation filter and SVM did better using SVM. Using PSO, metaheuristics was better than Relief and equivalent concerning correlation filter.

For the **GAC** approach, as well as the in the GREYC database, the results when compared to GAIP, PSO, GA [24] and GA with new settings, performed poorly in all scenarios for all classifiers.

For both databases, the best results happened with GAIP using the two filters, Correlation and Relief, in terms of quantity of attributes, however the classification accuracy is better for the wrapper approach using the PSO metaheuristic. Regarding the computational time, it is costly with the use of the wrapper approach (GA and PSO), however in the hybrid approach the time was higher due to the filters used, and the performance is related to the database used, as occurred with the use of the database [10].

Among the proposed approaches to feature selection, the GAC approach presents much lower results, the hypothesis raised is that, as in this approach, the filters are applied in the crossover step, which can generate the presence of twin children, however, they can stop be twins through a probabilistic rate when reached, which can improve the accuracy for classification algorithms enough.

6 Conclusion

As already discussed, the technic of feature selection is one of the most fundamental steps in pattern recognition, data mining, and machine learning and with it, dataset containing many features may need high computational power. Thereby, the selection of a subset of features relevant may increase the accuracy of the classification.

This work has shown a comparative analysis between two methods of feature selection using a wrapper and two hybrid approaches with genetic algorithms in databases of typing dynamics for gender recognition. Where our results showed that for databases with characteristics similar to [10], with many attributes and instances, the k-NN algorithm performs well even without selection, since the SVM algorithm after the feature selection using PSO having promising results. For databases with characteristics similar to [31], with fewer attributes and fewer instances, the SVM algorithm performs well without selection, after the feature selection using the PSO, k-NN presenting promising results.

Acknowledgments

This work was supported by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES). This research was supported by NPAD/UFRN.

References

[1] ANANYA and S. SINGH. Keystroke dynamics for continuous authentication. In *Proceedings...*, pages 205–208. 8th International Conference on Cloud Computing, Data Science Engineering (Confluence), Jan 2018.

[2] M. Antal and G. Nemes. Gender recognition from mobile biometric data. In *Proceedings...*, pages 243–248. IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), May 2016.

[3] N. Bidi and Z. Elberrichi. Feature selection for text classification using genetic algorithms. In *Proceedings...*, pages 806–810. 8th International Conference on Modelling, Identification and Control (ICMIC), Nov 2016.

[4] E Chandra Blessie and E Karthikeyan. Sigmis: A feature selection algorithm using correlation based method. *Journal of Algorithms & Computational Technology*, 6(3), 2012.

[5] D. R. Cavalcante Bandeira, A. M. de Paula Canuto, M. Da Costa-Abreu, M. Fairhurst, C. Li, and D. S. Costa do Nascimento. Investigating the impact of combining handwritten signature and keyboard keystroke dynamics for gender prediction. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 126–131, Oct 2019.

[6] A. Darabseh and A. S. Namin. Effective user authentications using keystroke dynamics based on feature selections. In *Proceedings...*, pages 307–312. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Dec 2015.

[7] C. Darwin. *Origem das espécies*. MADRAS, 1994.

[8] K. Faceli, A. C. Lorena, J.Gama, and A. C. P. L. F. Carvalho. *Artificial Intelligence: A Machine Learning Approach*. LTC, Rio de Janeiro, 2011.

[9] M. Fairhurst and M. Da Costa-Abreu. Using keystroke dynamics for gender identification in social network environment. In *Proceedings...*, pages 1–6. 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), Nov 2011.

[10] R. Giot, M. El-Abed, and C. Rosenberger. Greyc keystroke: A benchmark for keystroke dynamics biometric systems. In *Proceedings...*, pages 1–6. "IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems", Sep. 2009.

[11] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11(1):10–18, 2009.

[12] Z. Jie, Y. Li, and R. Liu. Social network group identification based on local attribute community detection. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 443–447, 2019.

[13] H. Kaur and P. Khanna. Privacy preserving remote multi-server biometric authentication using cancelable biometrics and secret sharing. *Future Generation Computer Systems*, 102:30–41, 2020.

[14] A. Kawamura and B. Chakraborty. A hybrid approach for optimal feature subset selection with evolutionary algorithms. In *Proceedings...*, pages 564–568. IEEE 8th International Conference on Awareness Science and Technology (iCAST), Nov 2017.

[15] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.

[16] Junhong Kim, Haedong Kim, and Pilsung Kang. Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection. *Applied Soft Computing*, 62:1077 – 1087, 2018.

[17] F. Koumi, M. Aldasht, and H. Tamimi. Efficient feature selection using particle swarm optimization: A hybrid filters-wrapper approach. In *2019 10th International Conference on Information and Communication Systems (ICICS)*, pages 122–127, 2019.

[18] L. E. A. S. Santana. Metaheuristic feature selection, 2019. `https://github.com/lauraemmanuella/MetaheuristicsFeatureSelection`, Last accessed on 2020-03-15.

[19] Cheng Li. *Extending the Predictive Capabilities of Hand-oriented Behavioural Biometric Systems*. PhD thesis, University of Kent,, September 2016.

[20] J. Li and H. Liu. Challenges of feature selection for big data analytics. *IEEE Intelligent Systems*, 32(2):9–15, Mar 2017.

[21] Julliana Marques, Tuany Nascimento, Brenda Vasiljevic, Laura Santana, and Márjory Costa-Abreu. An investigation of the predictability of the brazilian three-modal hand-based behavioural biometric: a feature selection and feature-fusion approach. *Journal of the Brazilian Computer Society*, 26, 12 2020.

[22] B. V. S. Mendes. Analysis of feature selection on the performance of multimodal keystroke dynamics biometric systems. Technical report, Universidade Federal do Rio Grande do Norte, Natal, September 2017.

[23] Tuany Mariah Lima do Nascimento. featureselectionhybrid-approach, 2019. `https://github.com/tuanymariiah/featureSelectionHybrid-Approach/tree/desenvolvimento`, Last accessed on 2020-03-15.

[24] Tuany Mariah Lima do Nascimento, Andrelyne Vitória Monteiro de Oliveira, Marjory Da Costa Abreu, and Laura Emmanuella Alves dos Santos Santana Santana. An investigation of genetic algorithm-based feature selection techniques applied to keystroke dynamics biometrics. 2019.

[25] Xiuqin Pan, Limiao Xue, Yong Lu, and Na Sun. Hybrid particle swarm optimization with simulated annealing. *Multimedia Tools and Applications*, 78(21):29921–29936, 2019.

[26] U. Revathy, Dr. M. Karnan, and Dr. S. Meenakshi Sundaram. Personal identification systems for keystroke dynamics with genetic algorithm and particle swarm optimization. 2017.

[27] B. S. Saini, N. Kaur, and K. S. Bhatia. Performance comparison of particle swarm optimization and genetic algorithm for feature subset selection in keystroke dynamics. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pages 684–689, 2019.

[28] L. E. A. S. Santana. *Otimizaçãõ em comitês de classificadores :*. PhD thesis, Universidade Federal do Rio Grande do Norte, Natal, 2012.

[29] L. E. A. S. Santana and A. M. P. Canuto. Filter-based optimization techniques for selection of feature subsets in ensemble systems. *Expert Systems with Applications*, 41(4, Part 2):1622 – 1631, 2014.

[30] J. Shen, J. Xia, X. Zhang, and W. Jia. Sliding block-based hybrid feature subset selection in network traffic. *IEEE Access*, 5:18179–18186, 2017.

[31] V. R. Da Silva, J. C. G. De Araujo Silva, and M. Da Costa-Abreu. A new brazilian hand-based behavioural biometrics database: data collection and analysis. In *Proceedings...*, pages 1–6. 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), Nov 2016.

[32] Ioannis Tsimperidis, Avi Arampatzis, and Alexandros Karakos. Keystroke dynamics features for gender recognition. *Digital Investigation*, 24:4 – 10, 2018.

[33] Chandrasekar Venko and S Kumar. Biometric based keystroke dynamics authentication - a review. *Asian Journal of Research in Social Sciences and Humanities*, 6:698, 01 2016.