User Authentification through Keystroke dynamics based on ensemble learning approach

Benoît Martin Azanguezet Quimatio, Olive Flore Yatio Njike, Marcellin Nkenlifack

User Authentification through Keystroke dynamics based on ensemble learning approach

Benoît Martin Azanguezet Quimatio[1], Olive Flore Yatio Njike[2], Marcellin Nkenlifack[3]

[1]University of Dschang, CAMEROON
[2]University of Dschang, CAMEROON
[3]University of Dschang, CAMEROON

*E-mail : azanguezet@gmail.com, yatioolive93@gmail.com, marcellin.nkenlifack@gmail.com

Abstract

Today, we live in a ubiquitous networked world, with the ability to connect to network systems regardless of time and place, using a variety of information technology devices. This increase in connectivity has raised concerns about the security of systems in protecting users' personal information and private data. The password has been the most common solution for user authentication. But it becomes vulnerable when a third party acquires it illegally. In order to increase security during authentication, several methods have been used such as keystroke dynamics which is a behavioral biometric. Previous works have demonstrated the feasibility of user authentication using keystroke dynamics. In this research, we propose an authentication method based on Bagging classifier set. This method is composed of three Bagging ensembles formed by SVM, KNN and decision tree classifiers. The outputs of these three ensembles are then merged and a majority vote was applied to obtain the final result. The proposed approach was evaluated using the CMU dataset. The final result obtained by our approach is promising, achieving an accuracy of 95.65%.

Keywords

Keystroke dynamics, classifier, Bagging, User authentication

I INTRODUCTION

With the emergence of telecommunications and competition, data security and user identification have become paramount. Computer security is the set of technical, organizational, legal and human means aiming at preventing the unauthorized use, the misuse, modification or misappropriation of the information system. In order to fight against fraud and impostors, it is necessary to impose a secure authentication method of the user, and today many solutions are biometric.

Biometric is initially the science of "measuring living things". It can be considered as an attractive solution for user authentication: the relationship between the authentication factor (biometric data) and the user is very strong. In the last three decades, the history of decades, the history of biometrics has marked a turning point, with the development of several techniques. Biometric techniques can use different characteristics attached to individuals. They have been developed to automatically verify a person's identity Prabhakar, Pankanti, and Jain (2003). Biometric modalities (Idrus, Cherrier, Rosenberger, and Bours (2014)) can be divided into three main

classes including morphological modality related to body shape (retina, voice, fingerprints), behavioral modality related to a person's behavior (gait, signature dynamics, keystroke dynamics) and the biological modality linked to the interior part of a living organism (heartbeat, ADN, blood) (Yampolskiy and Govindaraju (2008)).

Keystroke dynamics (KD) is a behavioral biometrics technique that allows to recognize an individual by the way he or she types on a computer keyboard.To identify a person based on his typing style, several parameters such as the time of pressure on each key, the time of release, the time of flight between two keys or even the number of fingers were used in (Idrus et al. (2014)). This authentication system is not likely to disturb the user experience, since "the system is low cost, non-intrusive, permanent and weakly constrained". It presents some instabilities due to transient factors such as emotions, stress, drowsiness, and many others and also depends on external factors such as the keyboard input device used (Yampolskiy and Govindaraju (2008)).

Some studies in the field of typing dynamics have relied on unique classifiers to authenticate users as in Zheng and Elmaghraby (2021). On the other hand, the authors in Shekhawat and Bhatt (2022) have used the Random Forest ensemble method to identify people. Although these studies may have had an improvement on previous work, the results obtained are not significant as they obtained 93.1% and 85% respectively.

In this paper, we propose an authentication approach based on several sets Bagging classifiers in order to obtain better performances. The proposed method consists of three sets of heterogeneous Bagging classifiers. In each Bagging ensemble, we used as base classifiers SVM , KNN and decision tree. The results of these three ensembles are then merged and a majority vote is applied to obtain the final decision. The rest of this paper is organized as follows : section 2 presents a state of the art on keyboarding dynamics; section 3 presents the preliminaries we used in our approach; section 4 presents our approach; section 5 shows the experiments performed and the results obtained; the conclusion of this paper is made in section 6.

II LITERATURE REVIEW

The very first research on keystroke dynamics was conducted by Gaines, Lisowski, Press, and Shapiro (1980). Inspired by the idea that individuals have unique rhythms when typing on a computer keyboard. It was through this study that the concept of digraphs was born. A digraph is a pair of keystrokes and the time between the first and second keystrokes. This study had a 100% success rate but this result was insignificant said Gaines et al. (1980) because only 7 subjects participated in the study and a significant amount of adjustments on the metrics had to be made. Umphress and Williams (1985) conducted a more thorough experiment and gave more credibility to the idea that keyboard dynamics was viable.

Today, various problems related to information security, more precisely user authentication in different systems, are solved by means of advanced artificial intelligence methods. In this field, various studies are intensively developed (Foster, Koprowski, and Skufca (2014); Porwik and Doroz (2014); Trajdos and Kurzynski (2015); Giot, Dorizzi, and Rosenberger (2015)). Although a number of searches are based on long text (free) input Wesołowski, Porwik, and Doroz (2016), Porwik, Doroz, and Wesolowski (2021), it is difficult to compare results from individual studies due to differences in data format, number of people participating in the study. Pleva, Kiktova, Juhar, and Bours (2015) presents an approach for person identification using acoustic monitoring of the required word typing on the monitored keyboard. In the work (Alsultan and Warwick (2013)), the authors pointed out, among other things, that a higher probability of intru-

sion occurs when a longer pause in the activity of a legitimate user was detected. Wesołowski et al. (2016) proposed a user profiling method and an intrusion detection method based on machine learning algorithms. In order to improve the authentication of users by their keystroke dynamics, some researchers have used deep learning. Tewari (2022) presents a survey to identify the various works on deep learning. Sun et al. (2020) developed a fraud detection system based on continuous authentication (KOLLECTOR) by applying the concept of deep learning: GRUBRNN (Gated Recurrent Unit-Bidirectional Recurrent Neural Network) for cell phones. They obtained an accuracy of 94.24%.

III PRELIMINARIES

We describe in this section the different methods used in our approach.

3.1 SVM

The support vector machine (SVM) Ma and Guo (2014) is a kernel-based supervised learning algorithm that classifies data into two or more classes. The SVM classifier is able to find the optimal hyperplane that separates two classes. This optimal hyperplane is a linear decision boundary that separates the two classes while leaving the largest margin between the samples of the two classes.

3.2 K-NN

The k-NN (k-Nearest Neighbours) method (Hu, Gingrich, and Sentosa (2008)) or k-nearest neighbors is a standard classification algorithm which relies exclusively on the choice of the classification metric. It is non-parametric (only k must be fixed) and is based only on training data. The idea is the following: from a labeled database, we can estimate the class of a new data by looking at the majority class of the k closest neighboring data (hence the name of the algorithm). The only parameter to define is k, the number of neighbors to consider.

3.3 Decision tree

Decision trees are decision support tools and represent a set of choices in the graphic form of a tree. The different possible decisions are located at the ends of the branches (the "leaves" of the tree), and are reached according to the decisions made at each step. They have the advantage of being easy to read and quick to execute.

3.4 Bagging

Bagging was introduced by Quinlan et al. (1996). It is an acronym for Bootstrap Aggregating. It improves the accuracy of a classifier by generating a composite model that combines several classifiers, all derived from the same inducer (learning algorithm). It uses a voting approach that is implemented differently, in order to combine the results of the different classifiers. Each instance is chosen with equal probability.

3.5 Majority vote

Majority voting (Kokkinos and Margaritis (2014)) is a fusion method operating on the outputs of classifiers and also considered a combination method. In its combination scheme, the classification of an unlabeled instance is performed based on the class that gets the most votes (the most frequent vote). This method is also known as plural voting or the basic set method.

This approach has been frequently used as a combination method to compare newly proposed methods.

IV PROPOSED METHOD

Recent works (Shekhawat and Bhatt (2022), Zheng and Elmaghraby (2021)) show us a great importance of artificial intelligence algorithms for user authentication through keystroke dynamics. In our approach, we propose to use an ensemble of classifiers to obtain better results compared to the use of individual classifiers. The figure shows the proposed approach.

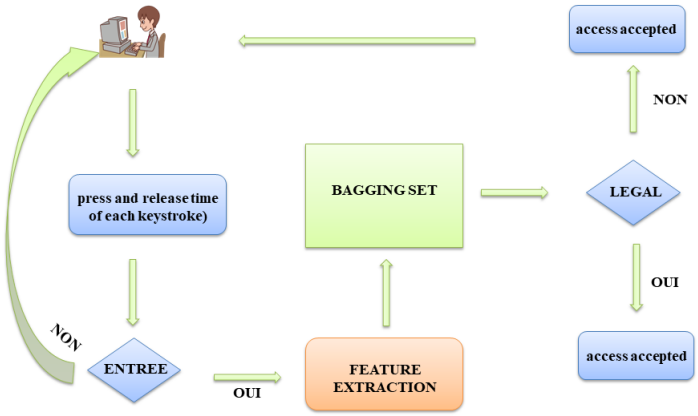

Figure 1: Figure describing our approach

4.1 Feature extraction

As we used a public database from Carnegie-Mellon University (CMU), three characteristics were considered in this study :

1. **Hold Time (HT)** : It is the time of pressing and releasing a given key. If we consider the E key, the hold time will be : E.KeyUp – E.KeyDown
2. **Down Down Time (DDT)** : This is the time that elapses between the pressing of a key and the next pressing of another key. Considering the E and D keys, the DDT will be : E.KeyDown – D.KeyDown
3. **Up Down Time (UDT)** : This is the time that elapses between the release of a key and the pressing of the next key. For two keys E and D, UDT will be : E.KeyUp – D.KeyDown

4.2 Classification

The proposed approach is mainly based on the Bagging ensemble method to identify users by their typing style with good performance. For this purpose, three sets of classifiers have been trained with the Bagging ensemble method as shown in figure 2. Each of them consists of three heterogeneous individual classifiers. The classifier ensembles work simultaneously and are trained on the data reserved for training. Thus, three different outputs are obtained for these three classification processes. The outputs of these three Bagging ensembles were merged to determine the final decision. This merging was done by applying majority voting. As a base classifier, we used the SVM classifier, the KNN classifier and the Decision Tree classifier.

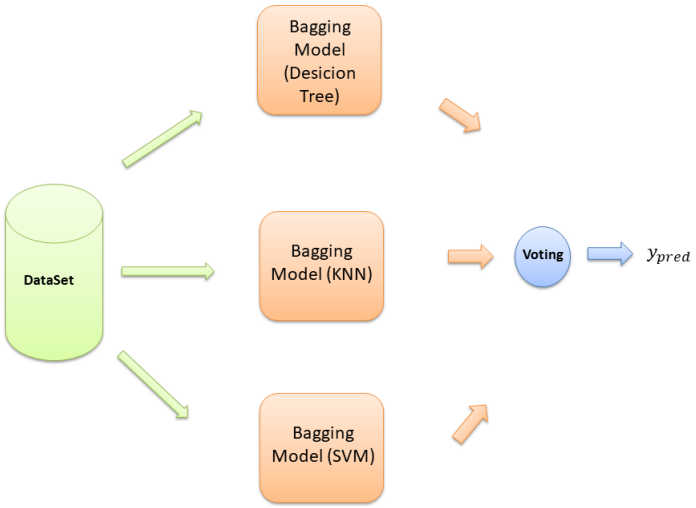

Figure 2: Bagging model

V EXPERIMENTATION AND RESULTS

5.1 Dataset

In this experiment, The Carnegie-Mellon University (CMU) fixed-text data set is used Killourhy and Maxion (2009). This dataset consists of three types of timing information namely hold time, key pressed time and key pressed and released time. This data was collected from 51 users and each user was asked to enter the password **.tie5Roanl**. Each user typed this password in 8 different sessions with 50 repetitions on each session. In total, each user typed the password 400 times. Between sessions, the user had to wait at least one day. There are a total of 31 features for each user trial, 11 of which are hold times ending with the input key, 10 of which are key press times, and 10 of which are key press times.

5.2 Data exploration

In the CMU database, there are 31 features as described in 5.1 including 11 features for **Hold Time**, 10 features for **Down Down Time** and 10 features for Up Down Time. Our analysis is based on the keystroke data of 6 users selected from the 51 users. As shown in figures 3, 4. They show the results of this analysis.

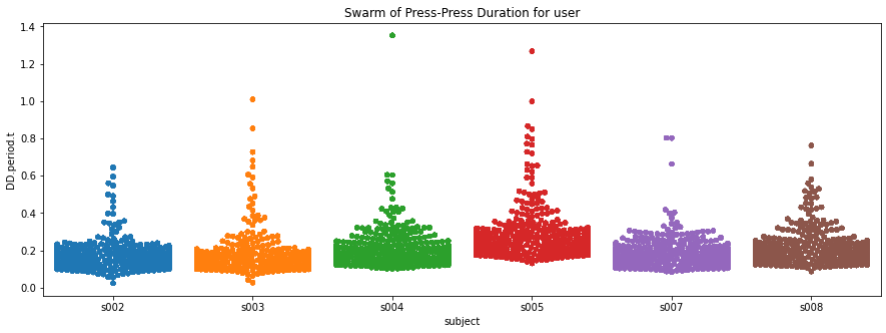

Figure 3: Graph of next and previous key press times (DD) between 6 users

Figure 3 shows the variation of the time between pressing the **period** key and the **t** key for 6 users. According to the figure, the user **s004** takes much more time between pressing the two keys, then the user **s005** and the user **s002** is the one who takes less time to press these two keys.

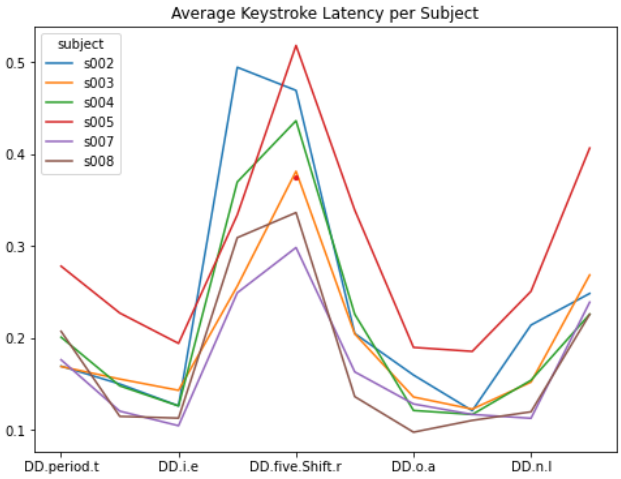

Figure 4: Variation of pressure times between two consecutive keys (DD) of several characters of 6 users

Figure 4 shows the time curve (DD) between two consecutive key presses of 6 users. The observation made on these curves shows that the user **s005** puts much more time between the interval of pressure of a key and the pressure of the following key and the user **s007** is the one who puts less time.

5.3 Performance analysis

We are used three performance evaluation metrics to evaluate our results : Accuracy , Precision and recall.

Accuracy: It is the number of samples correctly classified divided by the total number of classifications. More formally, accuracy is calculated as

$$\textbf{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision: it is the ratio between the number of correctly identified positive samples (corresponding to the right individual) in a given class and the number of correctly and incorrectly classified characteristics in this class. The formula is given by:

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

Recall : is the the ratio of correctly predicted positive to the total number of actually positive observations. The formula is given by:

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

5.4 Experimental classification

This phase of the proposed model shown in Figure 2 is studied in this experiment. In this phase, three different sets of classifiers are trained using Bagging. In the first set, a bagging with SVM as the base classifier is trained. In the second set, the KNN base classifier is used; and in the third set, the decision tree classifier is used. The results produced by the three bagging classifiers were merged and a majority vote was applied to obtain the final decision. In this phase, the data was divided into two, one for training and the other for testing. We used 80% of the data for the training of our model and 20% were reserved for the test. Since there are 400 hits for each user in the CMU database, 320 hits were used for training and 80 hits for testing. In the test phase, the test samples were classified using the Bagging patterns obtained in the training phase. We applied a normalization process on the data before using them for classification. Moreover, the boostrapping sampling technique is used in our method because it consists in replacing after each draw the data that have been selected in the data set. The advantage of this technique is to diversify the dataset to not have the same data on each set

The table 1 and the figure 5 on the other hand, show us the results obtained for each bagging ensemble method and the fusion of these ensembles. Finally figure 6 compares our approach with previous work (Shekhawat and Bhatt (2022), Zheng and Elmaghraby (2021)).

Model	Accuracy	Precision	Recall
Bagging SVM	90,20%	90%	90.20%
Bagging Decision Tree	92,80%	92.81%	92.79%
Bagging KNN	94,45%	94.5%	94.45%
Bagging Final	95,65%	95.67%	95.66%

Table 1: Result of each bagging method and the merging of bagging methods

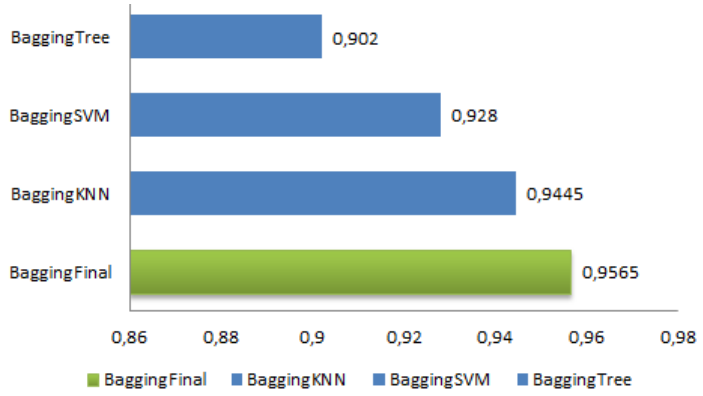

Figure 5: Graph of bagging model scores

VI SUMMARY AND DISCUSSIONS

The results obtained in table 1 or figure 5 show us the great importance of using ensemble methods. By applying our bagging ensemble method on each of these classifiers after hyper parameter settings, we have obtained 90.20%, 92.80%, and 94.45% accuracy respectively for the bagging ensembles with svm, decision trees and knn. In the table 1 or the figure 5, the result obtained for the combination of all the bagging classifiers is 95.65%. By comparing

this result to the three bagging sets, we find that the result obtained is clearly superior. In order to evaluate the performance of our approach, we compared our results to previous works (Zheng and Elmaghraby (2021), Shekhawat and Bhatt (2022)). Figure 6 shows us that our model achieves a better accuracy of 95.65% thus showing a modest improvement over previous work.

Model	Accuracy	Precision	Recall
Our model	0.9565	0.9567	0.9566
Zheng and Elmaghraby (2021)	0.93	-	-
Shekhawat and Bhatt (2022)(Random Forest)	0.85	-	-
Shekhawat and Bhatt (2022)(SVM)	0.76	-	-

Table 2: Results of our model with previous works

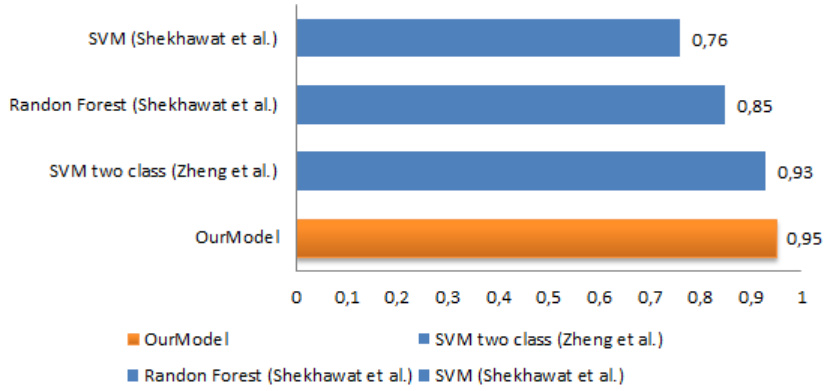

Figure 6: Score graph with other methods

VII CONCLUSION

In this paper, we presented our authentication approach based on the Bagging ensemble method to authenticate users by their way of typing and improve performance. This approach is the result of the fusion of three sets of classifiers using three basic classifiers namely: SVM, KNN and decision tree. It appears from the experiments carried out that the Bagging ensemble method with KNN as the base classifier provides a better result, followed by the Bagging ensemble with decision tree and Bagging with SVM. The combination of the three sets of classifiers and the use of majority voting to obtain the final decision of our approach gave us better accuracy, not only compared to the results of the different sets, but also compared to previous works. In order to improve the accuracy and stability of our model, we plan in our future work to study feature selection methods, combine this approach with mouse dynamics and finally use the deep learning approach.

REFERENCES

Alsultan, A., & Warwick, K. (2013). Keystroke dynamics authentication: a survey of free-text methods. *International Journal of Computer Science Issues (IJCSI)*, *10*(4), 1.

Foster, K. R., Koprowski, R., & Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomedical engineering online*, *13*(1), 1–9.

Gaines, R. S., Lisowski, W., Press, S. J., & Shapiro, N. (1980). *Authentication by keystroke timing: Some preliminary results* (Tech. Rep.). Rand Corp Santa Monica CA.

Giot, R., Dorizzi, B., & Rosenberger, C. (2015). A review on the public benchmark databases for static keystroke dynamics. *Computers & Security*, *55*, 46–61.

Hu, J., Gingrich, D., & Sentosa, A. (2008). A k-nearest neighbor approach for user authentication through biometric keystroke dynamics. In *2008 ieee international conference on communications* (pp. 1556–1560).

Idrus, S. Z. S., Cherrier, E., Rosenberger, C., & Bours, P. (2014). Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords. *Computers & Security*, *45*, 147–155.

Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In *2009 ieee/ifip international conference on dependable systems & networks* (pp. 125–134).

Kokkinos, Y., & Margaritis, K. G. (2014). Breaking ties of plurality voting in ensembles of distributed neural network classifiers using soft max accumulations. In *Ifip international conference on artificial intelligence applications and innovations* (pp. 20–28).

Ma, Y., & Guo, G. (2014). *Support vector machines applications* (Vol. 649). Springer.

Pleva, M., Kiktova, E., Juhar, J., & Bours, P. (2015). Acoustical user identification based on mfcc analysis of keystrokes. *Advances in Electrical and Electronic Engineering*, *13*(4), 309–313.

Porwik, P., & Doroz, R. (2014). Self-adaptive biometric classifier working on the reduced dataset. In *International conference on hybrid artificial intelligence systems* (pp. 377–388).

Porwik, P., Doroz, R., & Wesolowski, T. E. (2021). Dynamic keystroke pattern analysis and classifiers with competence for user recognition. *Applied Soft Computing*, *99*, 106902.

Prabhakar, S., Pankanti, S., & Jain, A. K. (2003). Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, *1*(2), 33–42.

Quinlan, J. R., et al. (1996). Bagging, boosting, and c4. 5. In *Aaai/iaai, vol. 1* (pp. 725–730).

Shekhawat, K., & Bhatt, D. P. (2022). Machine learning techniques for keystroke dynamics. In *Proceedings of data analytics and management* (pp. 217–227). Springer.

Sun, L., Cao, B., Wang, J., Srisa-an, W., Philip, S. Y., Leow, A. D., & Checkoway, S. (2020). Kollector: Detecting fraudulent activities on mobile devices using deep learning. *IEEE Transactions on Mobile Computing*, *20*(4), 1465–1476.

Tewari, A. (2022). Keystroke dynamics based recognition systems using deep learning: A survey.

Trajdos, P., & Kurzynski, M. (2015). An extension of multi-label binary relevance models based on randomized reference classifier and local fuzzy confusion matrix. In *International conference on intelligent data engineering and automated learning* (pp. 69–76).

Umphress, D., & Williams, G. (1985). Identity verification through keyboard characteristics. *International journal of man-machine studies*, *23*(3), 263–273.

Wesołowski, T. E., Porwik, P., & Doroz, R. (2016). Electronic health record security based on ensemble classification of keystroke dynamics. *Applied Artificial Intelligence*, *30*(6), 521–540.

Yampolskiy, R. V., & Govindaraju, V. (2008). Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, *1*(1), 81–113.

Zheng, Y., & Elmaghraby, A. S. (2021). Cybersecurity enhancement using recurrent neural networks and keystroke dynamics. In *Multimodal image exploitation and learning 2021* (Vol. 11734, p. 117340D).