A Clinical Decision Support System to Help the Interpretation of Laboratory Results and to Elaborate a Clinical Diagnosis in Blood Coagulation Domain

Francois Lasson[2], Alban Delamarre[1,2], Pascal Redou[3], and Cedric Buche[2]

[1] VISAGE Laboratory, FIU, USA,
adela177@fiu.edu
[2] LAB-STICC, ENIB, France,
{lasson, redou, buche}@enib.fr,
[3] LATIM, France

Abstract. Hemophilia is a rare hemorrhagic disorder caused by clotting factor deficiencies that leads to a less efficient coagulation system. Treatments of this pathology rely on a patient's subjective assessment which reflects a need for a laboratory assay able to predict the clinical patient phenotype. According to the literature, global assays such as thrombin generation (TG), are good predictors of bleeding episodes and therefore seem to be good candidates to fit this need. Nevertheless, the result of the TG assay, known as thrombogram, is difficult to interpret for non-expert clinicians. In this paper, we present a machine learning-based clinical decision support system which goal is to help clinical decision making. In doing so, we have adopted several approaches in order to evaluate well-known machine learning algorithms, in terms of accuracy and robustness, on a thrombogram database generated using numerical simulations. Obtained results, 95.57% of accuracy using a cascade of a SVM and MLPs to classify all categories and 98.10% of accuracy for the binary case hemophilia A/B, prove that our proposal can efficiently diagnose hemophilia.

Keywords: Machine Learning-Based Clinical Decision Support System, Thrombin Generation Assay, Hemophilia Diagnosis, Scikit-Learn

1 Introduction

Blood coagulation is a biological phenomenon leading to clot formation that prevents and stops bleeding after vascular damage. This complex system is regulated by a judicious equilibrium between the procoagulant pathways that are responsible for clot formation and the anticoagulant pathways that regulate and inhibit this process. An imbalance in this equilibrium may cause two kinds of pathologies: thrombotic disorders and hemorrhagic disorders. Hemophilia is a hemorrhagic disorder caused by deficient clotting factors that leads to a less efficient coagulation system. The main treatment for this pathology is replacement therapy that consists of clotting factor concentrate administrations. There are well-known laboratory assays that quantify the concentrations of these proteins in plasma but most of them are bad predictors of a bleeding episode. Hence, current treatments of hemophilia rely on a patient's subjective assessment using

physiological parameters such as a persistence of pain or a decrease of joint mobility. According to the literature, in particular [24], a global assay called Thrombin Generation (TG) seems to be a good candidate to predict patient phenotype. The result of this test, the thrombogram, is a curve that plots thrombin (the key enzyme of the blood clotting system) concentration over time. Although biologists have identified several discriminating features from these curves (eg. lag time, time to peak, peak or endogenous thrombin potential), this result is difficult to interpret for non-expert clinicians. To deal with this issue, Clinical Decision Supports (CDS) can be used. The goal of these systems is to help clinical decision making in order to increase the quality of care, the health outcomes but also to improve the cost-benefit by avoiding, in the hemophilia case, any overdose of clotting factor concentrates. A clinical decision can be diagnostic elaboration and also therapeutic orientation. In the context of hemophilia, a model able to individualize therapy (choice of the drug and optimization of its dose) will be really helpful. We present in this paper a machine learning based CDS system, which goal is to diagnose hemophilia from thrombin generation curves. Although there is no added clinical value of hemophilia diagnosis through TG, this work can nevertheless be seen as a preliminary study. Its aim is to demonstrate the interest of such a system to assist clinicians in interpreting thrombograms, a first step toward a therapeutic orientation.

This model has to determine the type of hemophilia (A or B) and its severity (Mild, Moderate, or Severe). Given the machine learning context, efficient learning algorithm strategies need large data sets. In the clinical domain, obtaining a huge data set is a long, complex and expensive process. As a consequence, a numerical thrombin generation model has been used in this study to generate data [10],.

The next section presents the context of this study. Section 3 presents the state of the art of classification techniques and a summary of the different types of CDS systems. Section 4 details the approach used in our context. In Section 5, evaluation criteria are identified and described according to the context; classification results are then presented. Section 6 contains results analysis. We finally discuss the issues of this work and the directions for future work in section 7.

2 Context

2.1 Problem Specification

Since the analysis of coagulation curves is a very complicated task and can be source of serious consequences if misinterpreted, our objective is to create a CDS. Given a TG curve as an input, this system should be able to provide a complete diagnosis to clinicians, i.e. determining whether the patient is hemophiliac or not and if so, the type of hemophilia and its severity. Given a space $\mathcal{V}$ of unlabeled data and $\mathcal{Y}$ a finite set of labels, we have $\mathcal{X} = \mathcal{V} \times \mathcal{Y}$ the space of labeled samples. Let $\mathcal{D} = \{x_1, x_2, ..., x_n\}$ be a dataset composed of n labeled instances, where $x_i = \langle v_i \in \mathcal{V}, y_i \in \mathcal{Y} \rangle$ and v_i a vector representing a time series of length m such as $v_i = \{t_1, t_2, ..., t_m\}$. The objective is to find the best classifier $\mathcal{C}$ which for a given time series v associates a label y such as $\mathcal{C}(v) = y$ with $\langle v, y \rangle \in \mathcal{X}$.

2.2 Thrombogram Dataset

In the clinical context, the construction of a large database is a complex, long and expensive process. This is particularly true for the hemophilia context owing to the prevalence of this disorder. To deal with this difficulty, we have implement a numerical thrombin generation model in order to generate thrombograms using numerical simulations. As defined in [10], 41 biochemical reactions between 35 proteins are taken into account to construct a system of ordinary differential equations. Its resolution provides the thrombin concentration overtime which is a thrombogram. The simulation of a hemophilia patient simply consists of lowering factor VIII (hemophilia A) or factor IX (hemophilia B) initial concentration. Using artificially generated data provides two advantages in the clinical domain. As noted before, data acquisition is complicated and very expensive as it requires finding patients with the corresponding pathologies who are not undergoing treatment. Furthermore, a numerical model allows us to generate large amounts of data covering a wide range of thrombogram types.

Our dataset $\mathcal{D}$ is made of 7 categories labeled as $\mathcal{Y}$ = {*Healthy, Hemophiliac A mild, Hemophiliac A moderate, Hemophiliac A severe, Hemophiliac B mild, Hemophiliac B moderate, Hemophiliac B severe*} and it is composed of 14000 thrombograms with the following proportions $Quantity$ ={*5000, 1500, 1500, 1500, 1500, 1500, 1500*} where each thrombogram contains 181 points. The integration step used to generate these data is equal to 5 mHz. This dataset provides the ability to perform different types of classification : healthy or hemophiliac, hemophiliac A/B, hemophiliac severity and all these categories at once.

2.3 Overview of the Approach

The performance of this CDS is based on its ability to successfully classify thrombograms using machine learning techniques. Therefore, 6 well-known classification models have been evaluated in this study. These are presented in the next section. The different steps of our approach are briefly described below:

1. Firstly, we created a dataset that includes the different categories of hemophilia. A large amount of thrombograms were generated in order to realize an efficient training set for each classification method.
2. Secondly, in order to optimize classification performances of each model, we tuned their hyper-parameters. Because of their interdependence, search algorithms such as grid search and random search have been applied. In an attempt to improve results, we also reduce the dimensionality of each thrombogram using feature extraction techniques.
3. Next, we compared the different classification techniques based on established medical criteria: accuracy, precision, recall and False positive rate.
4. Since some methods perform better on a specific classification set, we developed a cascade classification technique, using a combination of binary classifiers to separate the different types of hemophilia step by step.

3 State of the Art

3.1 Clinical Decision Support

In the clinical domain, diagnosis mistakes can have disastrous consequences. A CDS is a system which goal is to advise clinicians during the process of decision making in order to reduce diagnostic errors. In this domain, two approaches are conceivable: Knowledge-based CDS and non knowledge-based CDS.

A knowledge-based CDS is composed of inference and association rules established by experts. Patients' data is fed into the system to produce a diagnosis suggestion. `MYCIN` [21] for example, is a system composed of about 600 rules able to detect severe infections like bacteremia or meningitis. Systems like `Arden Syntax` [19] `GLIF3` [4] `PROforma` [22] were developed to allow health professionals to directly build their own CDS systems. According to the litterature, there is a plurality of application scope of this kind of CDS. However, in the domain of coagulation, even though it exists features regarded as the most discriminant by experts, the interpretation of thrombogram and therefore the detection of hemophilia and its characteristics remains a challenging task.

Non knowledge-based CDS systems, for their part, use artificial intelligence techniques to associate patterns in the data with pathologies. For instance, Shin et al. created a system able to detect cancer based on mass spectrometry and machine learning algorithms [20]. Due to the complexity of the thrombogram analysis, application of artificial intelligence techniques on TG curves seems to be a good option [24]. Therefore, we directed our choice towards non knowledge-based CDS.

3.2 Classification Methods

We want to classify time series using a static approach since all of the TG curves of our database have the same length. Therefore, an assortment of supervised parametric and non-parametric techniques can be used for comparison, such as well-known Support Vector Machines, `SVM` [18] [8] [26] or K-Nearest Neighbors, `KNN` [6] which use distance functions to discriminate categories. Moreover, using a large dataset we can also correctly perform neural network training using a MultiLayer Perceptron `MLP` [13] [25]. Although this model is not recent, it forms the basis of the very popular deep learning techniques. In addition, for the purposes of completeness, we also performed a linear discriminant analysis `LDA` and applied the `Adaboost` [2] and the `Decision tree` [16] algorithms. All the classification results presented in this paper were obtained using the Scikit-Learn framework.

3.3 Extraction Methods

Extracting features from thrombograms and thus, reduce their dimensionality, could improve classification results. Therefore, we identify several feature extraction techniques. Piecewise Linear Approximation `PLA` [9], for example, applies the last square method on segments of the data. The Symbolic Aggregate approXimation algorithm combined with a Piecewise Trend Approximation, `SAX/PTA` [7] translates time series into strings, the `SAX` method is based on values whereas the `PTA` is based on slope.

Thrombograms can also be represented by coefficients, using Discrete Fourier Transform, `DFT` [1,23] and Discrete Wavelet Transform, `DWT` [5,15]. The obtained coefficients contain information in both the temporal and the frequency domain. To perform these feature extractions, we use the Pyts Python package.

4 Workflow

This section describes in detail the steps of classification performance optimization.

4.1 Hyper-parameter Search

In order to optimize classification performance for each model, we have tuned their hyper-parameters using search algorithms. In other words, for each type of classification, we have optimized the penalty parameter, the kernel function and its associated coefficients of the SVM classifier. We also searched for the optimal number of layers, the number of neurons in each hidden layers, the learning rate, the random seed, the activation function and the solver algorithm for a shallow MLP. In the case of the Decision Tree, we have tuned its maximum depth, its maximal number of features and the minimum number of samples needed to split an internal node or to create a new leaf node. Moreover, we have taken into consideration the number of nearest neighbors and the distance metric of the KNN, the number of estimators and the learning rate of the Adaboost classifier and the solver of the LDA. Given the limited number of these meta-variables, classical search algorithms are sufficient [12]. A grid search requires fixing all hyper-parameters to a given value except one, which varies across a finite set of values. A random search consists in setting all hyper-parameters to random values chosen in an established range. We initially used a grid search with a logarithmic scale, in order to query a wider range of values for each hyper-parameter, and to determine a subset of values. Next, we aimed to reduce this subset by using a second grid search with a linear scale. Then, we finally tested a large number of random hyper-parameter values within this subset. This process is used to gradually reduce the subset of possible hyper-parameter values and allows the system to find one of the best configurations.

4.2 Classification

Thrombograms are fed into the classification algorithms which, in turn, output a class label (healthy, hemophiliac etc.). As is common in practice, we carry out a k-fold cross-validation (CV) to train, optimize and evaluate these supervised models. First of all, the dataset is divided into two parts, taking into consideration the proportions of the label values. The first one is used for the training phase and the second one for testing. In order to tune each hyper-parameter, the training sub-dataset is then randomly split into k folds of approximately equal size. A rotation principle in which, $k-1$ of these folds are used to train the classifier while the remaining fold is used to validate its performance, is repeated k times until each fold has been used as a validation dataset [11]. At the end of this process, when its set of optimal hyper-parameter values has been found, the classifier is trained on the whole training sub-dataset and evaluated on the remaining part of

the database. In the clinical context, due to the lack of data induced by the prevalence of hemophilia disorder, performing a k-fold cross-validation with a high k value could improve the classifier performances. However, the database used in this study is generated by our numerical model. Large amounts of data covering a wide range of thrombogram types can be simulated. Therefore, we hypothesized that carrying-out a complex cross-validation would be unnecessarily computationally expensive. Considering that point, we use 80% of our numerical database for the training phase and 20% for testing. Then, the cross-validation is performed with two different numbers of folds : $k \in \{3,10\}$. The obtained results are presented and discussed in section 5.

4.3 Features extraction

In order to improve classification, we realized a feature extraction as a preprocessing step using techniques identified during the state of the art. Each technique was applied on the thrombogram dataset to reduce their dimensionality and, as a consequence, facilitate the learning process. For the purposes of comparison, we also used 4 features regarded as the most discriminant by experts: Time to Peak, Peak, Lag-Time and Endogenous Thrombin Potential.

4.4 Cascade

A method can be used to identify all 7 categories in a single classification process. However, another approach consists in using a cascade of classification models. Some categories can be pulled out of the dataset by a specific classifier. For example, thrombograms of healthy patients can be extracted using a classifier trained on discriminating healthy patient from hemophiliac. On the remaining hemophiliac sub-dataset we can then separate hemophiliac A from B using another classifier and so on until all categories are isolated. With this principle, we can use the best method at each step of the cascade.

5 Evaluation

To measure the efficiency and the robustness of our system, given the clinical context, we need to take into account specific criteria, they are presented in this first subsection. Results obtained are shown in the second one.

5.1 Evaluation criteria

The main objective of a CDS is to help clinicians in the decisions, and thus to reduce medical errors. The worst possible case is the prediction of an absence of illness for an infected patient. No measures would therefore be taken to insure the safety of the patient. Our goal is thus to minimize these cases, they can be measured using the False Positive Rate (FPR). A second objective is to reduce the quantity of assays used for pathology detection. This goal is reached by detecting a majority of healthy patients : Recall. Finally, as we are working with artificially generated thrombograms, data are by definition clean of experimental noise. One last point is to study the noise robustness of the selected model.

5.2 Results

First of all, we have hypothesized in section 4, that carrying-out a complex cross-validation would be unnecessarily computationally expensive in this study. In order to test this assumption, we compared results obtained using cross-validation with two different numbers of folds : $k \in \{3,10\}$. To that end, an SVM classifier has been trained to discriminate the Hemophilia A/B case and tuned using a cross-validated grid search. On the same hardware, the 10-fold cross-validation took approximately 5 times longer than the 3-fold cross validation. Regarding the CV accuracies, k=10 leads to 98.0$\pm$0.4% and k=3 to 97.8$\pm$0.3%. Beyond the fact that these performances are very similar, the same optimal hyper-parameters values have been found by these two CV-grid search : polynomial kernel of degree 6 and a low penalty parameter equal to 5. Therefore, the last step which consists in training the classifier on the whole training sub-dataset leads to the same accuracy scores on the remaining part of the database : 98,1%. Hence, our assumption holds true.

Table 1 shows the result obtained with the different classification techniques using a 3-fold CV without feature extraction techniques. We can observe that `SVM` performs the best, above 94.49% for each type of classification. Moreover, it appears that the discrimination between hemophilia A and B is the most challenging one. Even thought *SVM* and *MLP* obtained high accuracies on this classification, all other methods failed to correctly discriminate these 2 categories. Regarding these MLPs, the hyper-parameter tuning phase results in shallow architectures composed of 3 hidden layers at most, optimal learning rates equal to $1e-3$, hyperbolic tangent activation function and 'LBFGS' solver.

Dataset	Method	Accuracy	Recall	Precision	F-Measure	FPR
All Categories	Decision Tree	75.00	68.16	67.94	67.97	32.06
	Adaboost	56.60	43.14	41.16	42.13	56.86
	KNN	78.95	72.68	72.53	72.51	27.47
	LDA	90.29	48.67	74.52	58.88	2.77
	SVM	**94.49**	**93.17**	**93.17**	**93.14**	**0.91**
	MLP	90.00	87.14	87.12	87.08	1.63
Healthy / Hemophiliac	Decision Tree	97.88	97.78	97.60	97.69	2.40
	Adaboost	88.07	87.42	86.95	87.17	13.05
	KNN	98.21	97.98	98.13	98.05	1.87
	LDA	96.73	97.09	94.04	95.54	3.45
	SVM	99.04	98.83	99.06	98.95	1.16
	MLP	**99.18**	**99.01**	**99.22**	**99.11**	**0.99**
Hemophilia A/B	Decision Tree	69.63	69.62	69.64	69.62	30.36
	Adaboost	52.85	52.86	52.87	52.81	47.13
	KNN	74.37	74.41	74.45	74.37	25.55
	LDA	57.53	57.31	57.69	57.50	42.24
	SVM	**98.1**	**98.1**	**98.1**	**98.1**	**1.9**
	MLP	90.83	90.83	90.83	90.83	9.17
Hemophilia Severity	Decision Tree	91.66	91.65	91.67	91.66	8.33
	Adaboost	84.04	83.87	84.14	84.01	16.13
	KNN	93.22	93.21	93.27	93.17	6.73
	LDA	92.95	89.48	89.38	89.43	5.31
	SVM	96.03	96.03	96.1	96.03	1.98
	MLP	**96.23**	**96.23**	**96.29**	**96.23**	**1.88**

Table 1. Averaged performance of each method without feature extraction techniques. Highest classification rate, Recall, Precision and F-measure and lowest FPR are shown in bold.

6 Discussion

In this section we compare the performance of each classification method. We also analyse the impact of using feature extraction and cascade classification. Finally, according to the identified criteria we present the most accurate method.

6.1 Performance of hemophilia detection

The main objective of this study is to detect hemophilia using thrombograms. Table 1 shows results obtained for the healthy and hemophiliac classification. `Decision Tree`, `KNN`, `LDA`, `SVM` and `MLP` can accurately discriminate thrombograms, their accuracies are above 96%. We can notice that `SVM` and `MLP` are slightly better for this classification. On the other hand, `Adaboost` performs relatively worst. In the performance evaluation section we pointed out that classifying a hemophiliac patient as healthy can have disastrous consequences. MLP outperforms other techniques regarding the *False Positive rate*, 0.99%. Moreover, in order to save hemophilia detection tests, we want to identify a majority of healthy patients, the best *Recall* is also obtained by `MLP`, 99.01%. Thus, a CDS based on a `MLP` is able to fulfil the two main criteria i.e., avoiding clinical errors and reducing costs.

6.2 Classification comparison after feature extraction

All classification techniques were also tested on extracted features in order to improve their performances. However, classification results using these features are less accurate than the ones obtained using whole thrombograms. Regarding the thrombogram database reduced to the 4 features identified by experts, a SVM obtains 85.81% of accuracy to classify All Categories, 98.83% of accuracy for the binary case Healthy/Hemophiliac, 83.62% for the Hemophilia A/B case and 95.33% for the Severity. In other words, this kind of dimensionality reduction results in a 8.7% decrease of accuracy for the All Categories classifier, 14.5% for the A/B classifier but only around 1% for the Healthy/Hemophiliac and the Severity classifiers. Results obtained by other techniques such as `PLA`, `SAX/PTA`, `DFT` and `DWT` are very similar. These lower results can be explained by the loss information induce by feature extraction techniques, particularly in the most challenging case A/B where the whole TG curve seems to be relevant.

6.3 Classification comparison using cascade classification

The cascade classification discriminates thrombograms categories by categories, healthy and hemophiliac first, hemophiliac A/B next, and finally the hemophilia severity. This technique allows to use the best classifiers for each classification tasks. We can notice a major contribution brought by the cascade technique. First of all, it allows to divide the problem and to identify where classification performances were not satisfying. Secondly, we incremented the number of classifier required by the system by decreasing the classification complexity, i.e. the cascade uses binary classifiers, except for the severity which could also be reduced to a combination of binary classifiers. However, this kind of approach also have a drawback. In fact, sets of instances received by

classifiers in the lower levels of the cascade already contains some misclassified thrombograms, thus errors made early in the process increase error rates of the following steps. In comparison to a single multiclass-SVM, the global performance achieved by a cascade composed of the best classifiers : a SVM for the A/B case and two MLPs for the others, is equal to 95.57%. In other words, this cascade principle results in a 1.08% increase of accuracy.

6.4 Complete diagnosis performance

In order to provide a complete diagnosis to clinicians specifying the type and the severity of the hemophilia, we consider results obtained for the All Categories classification with a cascade composed of a SVM and two MLPs. It achieves 95.57% of accuracy on this classification. We can therefore create a CDS able to provide a complete accurate diagnosis to clinicians.

6.5 Robustness

As mentioned in the previous sections, the database used in this study is composed of artificially generated thrombograms. Hence, data are clean of noise and are bound to be different than experimental ones. In order to approach an actual application, we added a Gaussian noise to our data with the variance $var(X) = k$ in which X is a centered Gaussian distribution and $k \in \{0, 5, 10, 15, 20\}$. It should be noted that the variance of this kind of experimental data usually seems to be in the range $[5, 10]$. To determine influence of this external parameter, some metrics were computed using noised datasets for the training, tuning and testing phases of these models. We focused on the two most challenging cases : Hemophilia A/B and All Categories and therefore, we used *SVM* classifiers. Given the fact that hyper-parameters are closely linked to the type of data used, we applied the process of hyper-parameter search mentioned in the previous section for each level of noise. Figure 1 shows that the addition of noise strongly degrades the entire model performance. Obviously, noise addition using $var = 20$ resulted in a 23.8% decrease of accuracy for the All Categories classifier, 34.7% for the A/B classifier. The poor performances induced by the noise addition can be explained using the table 1. All used algorithms accurately classify Healthy/Hemophiliac and severity types because thrombograms are really different for these categories. Yet, it's not the case for hemophiliac A/B which is more complex to discriminate. In fact, noise addition increases similarities between thrombograms, particularly for thrombograms which have weak amplitude (severe hemophilia).

These results point out the well-known difficulty of *SVM* to work on raw data, without preprocessing step [14]. To go beyond this limit, we decided to smooth and fit noised data using a median filter and a Savitzky-Golay filter with a constant window size. We transposed these results in the table 2. Yet, due to the constant window size, this process alters the shape of the peak and thus, no classification improvement has been noted (Figure 2).

In a second experiment, we also computed learning curves in which the quantity of noised data used for the training phase varies from 1-13999. As we can see in the figure

1, noise addition does not affect the quantity of needed data for training since only the amplitude is impacted and not the convergence speed.

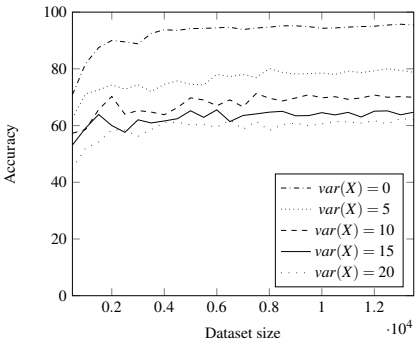

Fig. 1. The influence of noise on the hemophiliac A/B classifier performance

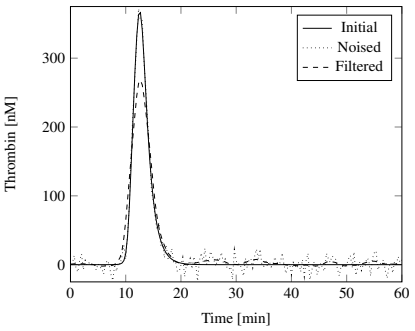

Fig. 2. Instance of a hemophiliac A mild illustrating the difference between raw, noised and filtered thrombograms

smooth & fit process	Variance	Accuracy	No. hemophiliac B classified as hemophiliac A	FPR	No. hemophiliac A classified as hemophiliac A	Recall
Without	0	98.1	17/900	1.9	883/900	98.1
	5	79.74	194/900	21.63	706/900	81.11
	10	72.15	244/900	27.19	656/900	71.48
	15	67.44	268/900	29.78	632/900	64.67
	20	63.44	333/900	37.04	567/900	63.93
With	0	95.22	40/900	4.52	860/900	94.96
	5	81.74	164/900	18.30	736/900	81.78
	10	71.22	266/900	29.56	634/900	72
	15	66.78	294/900	32.67	606/900	66.22
	20	61.78	325/900	36.22	575/900	59.78

Table 2. Noise robustness and application of a Savitzky-Golay and a median filter whose window sizes are $size = 9$ (P : Hemophiliac B, N : Hemophiliac A)

7 Conclusion

This final section highlights issues of a CDS to detect hemophilia and suggests ways for further research.

Hemophilia is a rare bleeding disorder that leads to a less efficient coagulation system. The main treatment for this pathology is a replacement therapy that consists of clotting factor concentrate administrations. Even though well-known biological assays that quantify the concentrations of these factors in plasma exist, there are not good predictors of bleeding episodes. Therefore, treatments rely on a patient's subjective assessment which reflects a need for a laboratory assay able to predict the clinical patient

phenotype. According to the literature, a global assays called TG appears as a good candidate to fit this need. However, its results are difficult to interpret for non-expert clinicians and that is why we suggested a non knowledge-based CDS. In this study, we only focussed on one of CDS application : the diagnostic elaboration. Given the obtained results, this work points out the benefit of such an approach.

A cascade composed of a `SVM` and two `MLPs` achieved the best results according to our evaluation criteria and provides an accurate global classification rate (95.57%). Moreover, it is able to accurately diagnose the hemophilia, its type and its severity, 99.18%, 98,1% and 96.23% respectively. Regarding the robustness of our model, adding a Gaussian noise strongly degrades the performance of the hemophiliac A/B classifier which could create an issue within an experimental context application. Moreover, as seen in the previous section, despite the theoretical interest of our noise filter, obtained results are disappointing.

For a first approach using machine learning techniques in the field of blood coagulation, few methods were used. Plenty of other well-known techniques could be tested on thrombograms. `HMM` [27] and `DTW` [17] generally perform well on time series and could appear as a great contribution to this work. Nevertheless, this kind of techniques have limitations like the SVM to deal with raw data. So, two options can be considered: (1) First of all, we could go more in depth during the smoothing process using a dynamic size window rather than a static one. This could potentially reduce the issue of peak alteration, and therefore increase the noise robustness of our model. (2) We could use deep learning methods. The main advantage of deep architectures, owing to their large number of hidden layers and the no-linearity associated to each ones, is their ability to extract highly abstract features from data. In addition, they can deal with very different types of data and can be applied to supervised but also unsupervised problems [3].

This study showed the ability of machine learning techniques to diagnose hemophilia. Obtained results open doors for other clinical application in the domain of blood illness, such as thrombophilia diagnosis and therapeutic orientation.

References

1. Agrawal, R., Faloutsos, C., and Swami, A. (1993). Efficient similarity search in sequence databases. *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms (FODO '93)*, 8958546:69–84.
2. Alonso, C. and Rodriguez, J. (2000). Time Series Classification by Boosting Interval Based literals. *Inteligencia Artificial*, 4(11):2–11.
3. Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2006). Greedy layer-wise training of deep networks. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, NIPS'06, pages 153–160, Cambridge, MA, USA. MIT Press.
4. Boxwala, A. A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q. T., Wang, D., Patel, V. L., Greenes, R. A., and Shortliffe, E. H. (2004). GLIF3: A representation format for sharable computer-interpretable clinical practice guidelines.
5. Chaovalit, P., Gangopadhyay, A., Karabatis, G., and Chen, Z. (2011). Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys*, 43(2):1–37.
6. Chaovalitwongse, W. A., Fan, Y.-J., and Sachdeo, R. C. (2007). On the Time Series K-Nearest Neighbor Classification of Abnormal Brain Activity. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 37(6):1005–1016.

7. Esmael, B., Arnaout, A., Fruhwirth, R. K., and Thonhauser, G. (2012). Multivariate time series classification by combining trend-based and value-based approximations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7336 LNCS(PART 4):392–403.
8. Gudmundsson, S., Runarsson, T. P., and Sigurdsson, S. (2008). Support vector machines and dynamic time warping for time series. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2772–2776.
9. Keogh, E., Chu, S., Hart, D., and Pazzani, M. (2003). Segmenting Time Series: A Survey and Novel Approach. *Data Mining in Time Series Databases*, pages 1–21.
10. Kerdelo (2006). *Méthodes informatiques pour l'expérimentation in virtuo de la cinétique biochimique. Application à la coagulation du sang. Modélisation et simulation.* Thèse, Université Rennes 1.
11. Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 14(12):1137–1143.
12. Mantovani, R., Rossi, A., Vanschoren, J., Bischl, B., and De Carvalho, A. (2015). Effectiveness of Random Search in SVM hyper-parameter tuning. *International Joint Conference on Neural Networks, IEEE*, pages 1–8.
13. Nanopoulos, A., Alcock, R., and Manolopoulos, Y. (2001). Feature-based classification of time-series data. *Information processing and Management*, 0056:49–61.
14. Nguyen, M. H. and de la Torre, F. (2010). Optimal feature selection for support vector machines. *Pattern Recogn.*, 43(3):584–591.
15. Popivanov, I. and Miller, R. (2002). Similarity search over time-series data using wavelets. In *Proceedings 18th International Conference on Data Engineering*, pages 212–221. IEEE Comput. Soc.
16. Quinlan, J. R. (1992). *C4.5: Programs for Machine Learning*, volume 1.
17. Ratanamahatana, C. and Keogh, E. (2004). Everything you know about dynamic time warping is wrong. *Third Workshop on Mining Temporal and Sequential Data*, pages 22–25.
18. Rüping, S. (2001). SVM Kernels for Time Series Analysis. *Time*, page 8.
19. Samwald, M., Fehre, K., de Bruin, J., and Adlassnig, K. P. (2012). The Arden Syntax standard for clinical decision support: Experiences and directions. *Journal of Biomedical Informatics*, 45(4):711–718.
20. Shin, H. and Markey, M. K. (2006). A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples.
21. Shortliffe, E. H., a.B. Lindberg, D., and E.H.S. (1976). *Computer-Based Medical Consultations: Mycin*. Elsevier.
22. Sutton, D. R. and Fox, J. (2003). The Syntax and Semantics of the PROforma Guideline Modeling Language. *Journal of the American Medical Informatics Association*, 10(5):433–443.
23. Wu, Y.-L., Agrawal, D., and El Abbadi, A. (2000). A Comparison of DFT and DWT Based Similarity Search in Time-series Databases. *Proceedings of the 9th International Conference on Information and Knowledge Management (CIKM)*, 35(2000-08):488–495.
24. Young, G., Sorensen, B., Dargaud, Y., Negrier, C., Brummel-Ziedins, K., and Key, N. (2013a). Thrombin generation and whole blood viscoelastic assays in the management of hemophilia: current state of art and future perspectives. In *Blood*, pages 1944–1950.
25. Yu, C. H., Bhatnagar, M., Hogen, R., Mao, D., Farzindar, A., and Dhanireddy, K. (2017). Anemic Status Prediction using Multilayer Perceptron Neural Network Model. 50:213–220.
26. Zhang, D., Zuo, W., Zhang, D., and Zhang, H. (2010). Time series classification using support vector machine with Gaussian elastic metric kernel. *Proceedings - International Conference on Pattern Recognition*, pages 29–32.
27. Zhang, Y. (2004). Prediction of Financial Time Series with Hidden Markov Models. *The Magazine Of The Fine Arts*, B. Eng(May):102.