Grammatical Evolutionary Techniques for Prompt Migraine Prediction*

Authors were omitted for blind review

ABSTRACT

The migraine disease is a chronic headache presenting symptomatic crisis that causes high economic costs to the national health services, and impacts negatively on the quality of life of the patients. Even if some patients can feel unspecific symptoms before the onset of the migraine, these only happen randomly and cannot predict the crisis precisely. In our work, we have proved how migraine crisis can be predicted with high accuracy from the physiological variables of the patients, acquired by a non-intrusive Wireless Body Sensor Network. In this paper, we derive alternative models for migraine prediction using Grammatical Evolution techniques. We obtain prediction horizons around 20 minutes, which are sufficient to advance the drug intake and avoid the symptomatic crisis. The robustness of the models with respect to sensor failures has also been tackled to allow the practical implementation in the ambulatory monitoring platform. The achieved models are non linear mathematical expressions with low computing overhead during the run-time execution in the wearable devices.

CCS Concepts

•**Computing methodologies** → **Genetic programming; Model verification and validation;** •**Computer systems organization** → Embedded hardware;

Keywords

Grammatical Evolution, prediction, migraine, biosignal

1. INTRODUCTION

The migraine is one of the most disabling neurological diseases. The migraine is a recursive and chronic headache presenting symptomatic crisis that leads to the degradation of the professional and social life of the migraine sufferers (migraineurs).

Stovner *et al.* in [22] show that the prevalence of the migraine in European population is around the 15%. Others, such us Lipton *et al.* in [14], or the WHO [26], limit this value to the 10% worldwide. The economic consequences of the migraine represent €1,222 per patient per year [13] in Europe, that means almost €125,000mill in the whole old continent. Most of the direct costs are due to absences from work or low job performance. When a migraine occurs, a cascade of neurological processes starts and leads to the pain. The most efficient way to stop this process and avoid the pain is to advance the intake of specific drugs. Therefore, the action mechanism of the medicine is able to block the symptoms before they appear. Thus, a prediction system becomes necessary.

As a chronic disease presenting symptomatic crisis, the migraine and the migraineurs are suitable for an ambulatory study. In the Internet of Things (IoT) era, dozens of commercial low power wireless monitoring devices allow making ambulatory monitorizations easier. These Wireless Body Sensor Networks (WBSNs) monitor biometric variables in a non intrusive way. The devices composing the WBSN are battery supplied and the capacity of their batteries is sufficient for short time operation, such as in mobile applications for sport training [12]. When an ambulatory monitorization is required, for example in medical alarms [1], despite the usage of low power microcontroller architectures or more efficient wireless communication interfaces, improving the availability of the monitoring devices becomes a challenge.

Longer battery operation means also less disruptions in the monitored data. This is critical in the context of ambulatory monitorization for medical alarms. In this way, several research groups work to reduce the consumption to enlarge the battery life. Complete low power monitoring systems have appeared such as the one of the SmartCardia company [23]. Others, for example, study the consumption in medical monitoring devices at the system level, such as Tobola *et al.* do in [24]. They also study the dependence of the battery life with the sampling rate of the biomedical signals in [25]. Works like Braojos *et al.* [7] analyze the technique of compressive sensing in biosignals to reduce the energy consumption of such monitoring devices. One mechanism to reduce the energy consumption of the monitoring devices and increase the operation time of the battery is to reduce the complexity of the processing in the embedded microprocessor, and hence the number of clock cycles required to execute the code. Our work also targets this goal, proposing modeling and prediction techniques with lower complexity than classical mechanisms, aiming a higher energy efficiency.

Some previous works have demonstrated that the intake of the medication in advance reduces the probability of pain. Hu *et al.* show in [11] that the pharmacokinetics of specific migraine treatments, such as some kinds of triptans, can abort the

*Information omitted for blind review.

GECCO'16, July 20–24, 2016, Denver, Colorado, USA.

DOI: 10.475/123_4

migraine in periods of time between 10 and 30 minutes before the pain starts. Previous works [3] have shown that the migraine prediction is possible in order to anticipate the intake of the drugs. The prediction of the migraine was faced by the authors using classical state-space modeling techniques. This paper presents a new approximation to the migraine prediction using Grammatical Evolution (GE), to compare the accuracy of this methodology against classic methods, reduce the model complexity and help on the automatic feature selection, and improve the robustness of previous approaches.

GE presents several characteristics that makes it suitable for our purpose of predicting migraine crisis. GE performs automatically feature engineering by means of symbolic regressions, as opposed to the time costly manual feature selection of the traditional methods [6]. This approach for a real world application has also the benefit of creating more efficient models in terms of energy. These models implement simple non-linear equations easily programmable in wireless monitoring devices.

Grammatical Evolution algorithms are Genetic Algorithms that can evolve complete programs encoding a set of pseudo random numbers on a chromosome to select the appropriate rules from a Backus Naur Form (BNF) grammar [20]. Genetic Algorithms and Genetic Programming, thanks to the improvement of computational resources have faced to many real world applications in different areas, such as prediction of economic events in financial applications, ecological simulation [2] or modeling of dynamic power consumption of servers [21] in data centers. Related to health problems, GE has been used for classification purposes such as fetal heart rate in [9] or to model glycemia in humans as shown in [10].

The remainder of this paper is as follows. The experimental set-up is shown in Section 2; where the methodology and the parameters used in our GE modeling technique is described as well. Section 3 shows the results obtained and their discussion. Finally, some conclusions of this work are drawn in Section 4.

2. EXPERIMENTAL SET-UP

2.1 Data

The work presented in this paper is based on the data acquired in a real scenario. Data have been gathered using a WBSN in star topology. Four biometric variables have been measured using two non-intrusive sensing motes: the i) PLUX-Wireless Biosignals [16] to acquire the skin temperature (TEMP), the skin conductance or electrodermal activity (EDA), and the ECG signals, and the ii) Nonin Onyx II [15] to acquire oxygen saturation (SpO2). These motes transmit the data to an Android smartphone via Bluetooth. The smartphone transmitted the data periodically to a cloud storage system. Patients are asked to fill the relative changes of an on-going migraine in the smartphone app. These points draw a curve which is normalized (0 to 100%) and modeled as two semi-Gaussian curves [3].

Data are processed offline. Firstly, biosignals are synchronized and the HR signal is computed from ECG; then, to repair disruptions in data, a Gaussian Process Machine Learning (GPML) is followed [19, 18]. All the details regarding the data processing are discussed in a previous work [3].

For this paper, in order to show the results, data from two female patients have been selected (Patient A and B). Patient A is a young patient suffering from migraines with aura and no medical treatment. The data from Patient A are 20 migraines (acquired in two experimental studies of almost one month each). Patient B is a middle aged patient suffering from migraines without aura and

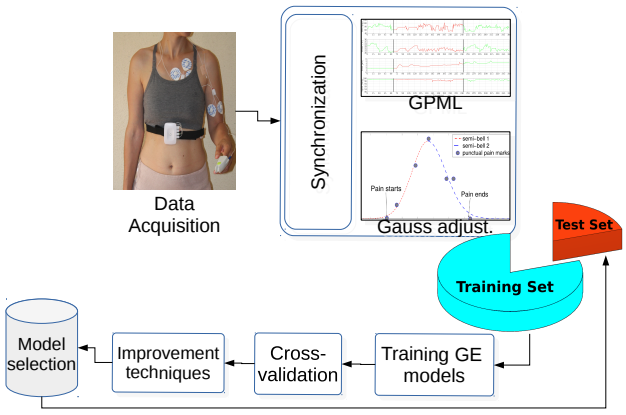

Figure 1: Overview of the followed methodology.

using preventive treatment. 13 migraines were acquired in one experiment during one month. The training dataset, M, is composed of 15 randomly selected migraines from Patient A and 8 randomly selected migraines from Patient B.

Despite the monitorization time is ideally 24 hours, most of the times this is not possible. Nevertheless, large periods of time around the migraine events have been chosen. With a period of time of 1 minute, the GE algorithms are fed with four hemodynamic inputs u_x, with $x = 1, 2, 3, 4$, to predict the Gaussian adjustment of the subjective pain marks. These inputs are the aforementioned: skin temperature, EDA, heart rate (HR) calculated from the ECG, and SpO2. The length of the input datasets range from 6 to 20 hours.

2.2 Methodology

The prediction of migraines has already been demonstrated by the authors in a previous work [3, 5]. In that case, classic state-space models were used to model the migraine, and a time-demanding basic study of feature dependence was conducted. In this paper, we present alternative models for migraine prediction using Grammatical Evolution (GE) techniques. According to our methodology in Figure 1, we focus the description of the training and the validation processes using GE to finally select the best models and perform the test. The remaining blocks are described in our previous work [3], but they are not required for a comprehensive understating of our current research.

GE [20] is a grammar-based form of Genetic Programming (GP) [17], used to generate programs in any language, where a Genetic Algorithm (GA) selects a group of production rules expressed in a Backus Naur Form (BNF) grammar previously defined. The GE leads to a mathematical expression which is a combination of features. This mathematical expression is an optimal solution obtained by applying Symbolic Regression (SR). The features are operations, combinations and transformations over the input signals (the biometric variables in our case). They are automatically selected by the GE and this is an advantage over the classic methods previously used to predict migraines.

GE, based on biology, evolves a population formed by a set of individuals. Each individual is a mathematical solution, and is represented by a chromosome. A chromosome is an array of integer numbers (genes), and each one of these defines the rule of the BNF to be applied. The solutions mutate and mix each other to create new ones in every generation. The input data are evaluated

```
N = { <Model>, <Op>, <PreOp>, <Input>, <Fcn>, <Const>,
      <Base>, <Exponent>, <Sign>, <Var>, <Pw>, <Fw> }
T = { +, -, *, /, Exp, Sin, Cos, Log, Avg, Sum, Drv,
      Max, Min, AvgFFT, DrvFFT, MaxFFT, MinFFT,
      0, 1, 2, 3,..., 130, TEMP, EDA, HR, SPO2, YR, YP }
S = <Model>

P = {
  I    <Model>      ::= (<Model><Op><Model>)
                     | <PreOp>(<Model>)
                     | <Fcn>(k-<Pw>,k-10,<Input>)
                     | <Var> | <Const>
  II   <Op>         ::= +|-|*|/
  III  <PreOp>      ::= Exp|Sin|Cos|Log
  IV   <Input>      ::= TEMP|EDA|HR|SPO2
  V    <Fcn>        ::= Avg|Sum|Drv|Max|Min
                     |AvgFFT|DrvFFT|MaxFFT|MinFFT
  VI   <Const>      ::= <Base>*Pow(10,<Sign><Exponent>)
  VII  <Base>       ::= 1|2|3...99
  VIII <Exponent>   ::= 1|2|3...9
  IX   <Sign>       ::= +|-
  X    <Var>        ::= TEMP(k-<Pw>)
                     | EDA(k-<Pw>)
                     | HR(k-<Pw>)
                     | SPO2(k-<Pw>)
                     | YR(k-<Pw>)
                     | YP(k-<Fw>)
  XI   <Pw>         ::= 10|11|12...130
  XII  <Fw>         ::= 0|1|2...9
}
```

Figure 2: BNF grammar used for migraine prediction with 10 minutes in advance. The inputs of the grammar are the hemodynamic variables: surface skin temperature (TEMP), electrodermal activity (EDA), heart rate (HR) and oxygen saturation (SPO2), as well as the past predicted pain level (YP).

Figure 3: Decoding process in GE that, starting from the genotype above, obtains the phenotype by mapping through the proposed grammar.

for each solution and the objective with the desired output is calculated. The best expression will survive with a higher probability in future generations.

A BNF grammar is represented by a set of parameters in the form $\{N, T, P, S\}$, where N is the set of non-terminals (coded symbols), T is the set of terminals (decoded expressions), P is the set of production rules to substitute the elements of N into T, and S is a non-terminal element of N used as starting symbol.

Given that our GE algorithm is generating predictive models, we have designed a grammar that produces phenotypes forming symbolic mathematical expressions for the target model.

Figure 2 shows, in BNF format, the version of the grammar that we have designed for the generation of predictive models. As seen in the figure, a predictive model (labeled as `Model` in Figure 2) is a combination of functions of six variables (terminals, T): the actual signal (`YR`), the set of inputs (the hemodynamic variables: `TEMP`, `EDA`, `HR` and `SPO2`), and the predicted signal (`YP`). There exist a future horizon, labeled as `Fw` and a past horizon, labeled as `Pw` in Figure 2.

One of the main benefits of GE is the capacity to create complex phenotypes from a genotype formed by integer values. In this regard, we show an example of the mapping process for the problem at hand using the grammar in Figure 2. Let's consider the genotype shown in the upper part of Figure 3 as a candidate solution. This individual is a genotype that represents a predictive model. However, it has to be decoded in order to obtain the symbolic expression that will define such model.

The mapping or decodification process is performed by applying the modulus operation between the current codon value and the number of options of each production rule corresponding to the non-terminal symbol being processed. Then, considering the example individual, the decoding process is performed as shown in Figure 3. It begins decoding the starting symbol, `<Model>`. Here, the first gene of the genotype is selected, 220. Given that the production for `<Model>` (rule `I`) has 4 different options, the selected value is option 0 (220 MOD 4 = 0), which corresponds to `<Model><Op><Model>`. Notice that the first option of the rule is indexed with 0, the second with 1 and so on. Next, the second gene is read and its value, 106, is used to decode the following non-terminal symbol, which is `<Model>` with production rule `I` once more. Then, after the modulus operation (106 MOD 4 = 2), the first `<Model>` is replaced by `<Fcn>(k-<Pw>,k-10,<Input>)`. Now, the following non-terminal symbol to be decoded is `<Fcn>`, whose production rule, `V`, has 9 different options. Then, the codon value, 91, is processed. Given that 91 MOD 9 = 1, the selected value is `Sum`. In the same way, the next codon, 123 for `<Pw>` is turned into 13 using rule `XI`, given that 123 MOD 120 = 3. The procedure continues until the full expression is (i) completely decoded or (ii) there are no codons to finish all the non-terminals, considering the chromosome invalid (sometimes, a wrapping process is used, but this is not our case). As a result, our genotype example returns the phenotype shown in the lower part of Figure 3, which represents a predictive model with past horizons of 10 and 13 units of time.

The grammar in Figure 2, and is the adaptation for 10 minutes of prediction horizon. This grammar is constituted by 12 production rules that are used to compose functions, variables and constants. The starting symbol, S, is the most general definition of an expression, the `<Model>` production rule. The others are non-terminals N production rules. In addition to the basic mathematical pre-operators `<PreOp>`, other basic mathematical functions are in `<Fcn>`. In these functions we have defined complex functions based on the Fast Fourier Transform (FFT) over the signals. In addition, the BNF considers to take past

Table 1: Parameters for the GE experiments.

Parameter	Value
Number of generations	150000
Population size	250
Probability of crossover	0.9
Probability of mutation	0.083
Chromosome length	100
Wrapping	No

samples of the input variables as described in the `Var` production rule. The maximum past horizon is 120 minutes (<Pw>), enough for our experiments. The output can also be based on the past predicted pain levels in a 10 minutes window backwards (<Fw>).

In this work we have used the JECO library published in [4] under GPL license. This work performs a mono-objective study, with the Normalized Root Mean Square Error (NRMSE), or fit, as objective function to minimize Eq. 1. The parameters used for training the models are shown in Table 1.

$$fit = 100 \times \left(1 - \frac{\|y - \hat{y}\|}{\|y - mean(y)\|}\right) \quad (1)$$

The JECO library compiles a Java code with the individuals (solutions) of each generation to evaluate them quickly. To reduce the number of compilations, the size of the population is high but still allows diversity of solutions in the next generations. The probability of mutation is the inverse of the number of rules, 12 in our case. Wrapping is not allowed, if a solution is not decoded it will return a non-valuable mathematical function. The length of the chromosomes is sufficient to avoid this situation in more than the 80% of the cases.

3. RESULTS

In this section we present the experimental results after applying our methodology based on GE for the prediction of migraines. First, the results of the training of the models are drawn, followed by the cross-validation results. After that, we present how these results can be benefited from the application of improvement techniques, and how the model selection phase is conducted. The dependency of the models with the inputs is studied, in order to select those suitable for implementation in a real monitoring device. With the selected models, a test is performed with the test dataset. Finally, an example of robustness against sensor failure is shown.

3.1 Training of models

Each model has been trained along 150000 generations with a solely objective: the minimization of the NRMSE. The four available hemodynamic variables i) skin temperature ($u_1[k]$), ii) electrodermal activity ($u_2[k]$), iii) heart rate ($u_3[k]$) and iv) oxygen saturation ($u_4[k]$) have been used as inputs. The output of the GE, $y_p[k]$ tries to fit the Gaussian adjustment of the real subjective pain, $y_r[k]$. Six different prediction horizons have been computed: from 10 to 60 minutes in step of 10 minutes, and no restrictions in the data selection have been imposed to the algorithm. 15 migraine events perform the training dataset for Patient A and 8 migraines for Patient B.

Regarding previous works, we hypothesize that models to predict migraines must be derived per patient, and thus, both patients are going to be trained separately. The training results for Patient A and B are shown in Figures 4a and 4b respectively. In

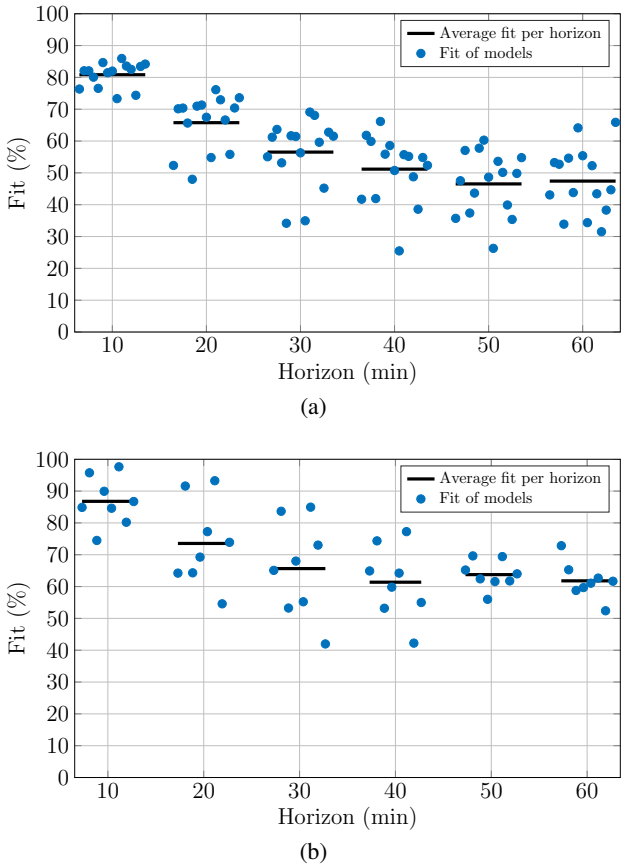

Figure 4: Training results. (a) Results for Patient A are good enough up to 20 minutes with an average fit of 65.8%; (b) Results for Patient B reach the 30 minutes of prediction with 65.6% of fit.

these figures each circle represents the fit achieved by each model, and for every prediction horizon. The horizontal black lines represent the average fit of all the models for every horizon. As expected, in average, lower fits are achieved for larger prediction horizons. In our criterion, a fit of 70% represents a good level of similarity and, with this threshold, the best results for Patient A are for 10 minutes of prediction horizon. The results for 20 minutes achieve a slightly lower fit (65.8% on average). Models trained for Patient B achieve better results, with average fits higher than 70% for 10 and 20 minutes of prediction and 65.6% of fit for 30 minutes.

In order to remove overfitted models, a cross-validation is going to be performed in the next section. In this way, we can remove those models overfitted in training.

3.2 Validation of models

In this section, we compute an M cross-fold validation to find those models that predict a migraine with higher accuracy for each prediction horizon. In this stage, predictions have been improved by means of repairing techniques. The authors have demonstrated in a previous work [5] how these techniques improve the fit, in average, in more than 10 points. We present an example of such improvement at the end of this section.

For each patient, a model M_i, with $i = 1, 2, \ldots, M$, trained with the $i-th$ migraine, is selected and validated with the remaining $M-1$ migraines in the training set. The average of the $M-1$ validations

Table 2: Validation results for Patient A.

# Model/ Horizon (min)	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
10	81.5 ±2.5	*NA*	74.2 ±9.0	78.2 ±6.7	**86.2 ±2.9**	**86.7 ±2.3**	77.1 ±6.1	72.5 ±12.6	77.9 ±6.7	**85.5 ±6.3**	80.5 ±5.7	80.2 ±5.8	84.2 ±4.3	73.4 ±9.4	78.1 ±9.8
20	44.4 ±39.8	58.4 ±15.2	**65.5 ±12.3**	**71.3 ±8.0**	58.7 ±11.2	59.1 ±15.8	58.4 ±13.1	56.9 ±17.1	53.3 ±0.7	64.2 ±15.7	**69.4 ±18.5**	60.9 ±12.3	54.3 ±21.8	60.5 ±14.4	64.7 ±17.0
30	66.9 ±0.4	44.7 ±22.2	52.6 ±7.9	45.4 ±19.8	23.4 ±12.8	44.5 ±20.1	43.5 ±20.2	43.1 ±21.5	73.5 ±0.0	**67.0 ±16.0**	30.2 ±21.1	54.9 ±14.2	64.6 ±2.6	**45.3 ±25.8**	**45.6 ±30.6**
40	24.4 ±2.5	52.8 ±15.8	35.0 ±18.6	42.6 ±10.1	78.2 ±0.0	48.6 ±11.1	39.7 ±7.9	39.0 ±0.0	*NA*	44.6 ±21.8	48.9 ±11.7	45.8 ±26.2	2.3 ±0.0	41.0 ±21.9	20.2 ±21.6
50	5.6 ±0.0	43.7 ±16.1	37.9 ±22.0	39.3 ±13.8	56.0 ±25.4	45.7 ±19.5	36.1 ±8.2	*NA*	20.4 ±0.0	40.4 ±8.7	46.9 ±22.4	50.0 ±6.3	*NA*	47.2 ±20.8	*NA*
60	22.4 ±11.1	31.1 ±1.6	30.7 ±19.2	*NA*	54.2 ±17.8	37.9 ±19.8	32.8 ±19.6	30.4 ±0.0	15.7 ±0.0	46.5 ±19.5	28.5 ±9.4	26.6 ±11.6	1.3 ±2.6	26.5 ±10.0	*NA*

Table 3: Validation results for Patient B.

# Model/ Horizon (min)	M1	M2	M3	M4	M5	M6	M7	M8
10	69.5 ±7.1	**87.2 ±10.3**	55.1 ±11.4	**85.8 ±5.7**	70.2 ±8.4	79.2 ±23.7	41.6 ±13.5	**91.7 ±5.9**
20	47.4 ±6.1	**72.2 ±19.9**	52.0 ±18.2	49.5 ±17.6	**76.6 ±7.7**	**78.2 ±13.6**	46.7 ±9.3	50.8 ±11.4
30	**74.1 ±0.0**	**60.1 ±11.7**	33.0 ±14.5	55.8 ±5.4	36.2 ±0.0	**70.3 ±10.9**	53.4 ±9.7	47.1 ±0.0
40	65.7 ±0.0	57.0 ±15.9	51.1 ±11.3	65.9 ±4.6	52.6 ±0.0	36.0 ±29.5	*NA*	28.9 ±14.8
50	45.8 ±0.0	32.3 ±49.8	*NA*	51.0 ±2.4	2.3 ±0.0	36.4 ±26.8	*NA*	*NA*
60	39.5 ±0.7	44.0 ±33.6	93.6 ±0.0	34.6 ±20.1	65.3 ±0.0	17.9 ±11.3	6.3 ±0.0	*NA*

is calculated after removing the overfitted validations (those with negative NRMSE), and the results are shown in Tables 2 and 3 for Patient A and B, respectively. These tables represent the average fit and also the standard deviation. Results with 0.0% of deviation represent models only able to validate 1 of the $M-1$ migraines in the training dataset. The higher the horizon, the lower the number of migraines correctly predicted. Notice that some models are not able to predict none of the migraines in the dataset (labeled as *NA*).

As expected from the training results, models are not able to reach more than 20 minutes with the defined level of accuracy (close to 70%). Thus, from this point, we discard prediction horizons greater than 30 minutes.

To avoid overfitting, 3 models have been selected to compute an average prediction. The models marked in bold in Tables 2 and 3 have been chosen as the best ones for each prediction horizon. The criteria have been as follows: i) first, to sort the models according to the number of migraines validated with a fit higher than 70%, then ii) to sort them by average fit and take the first three models. This may cause some undesired selections, such as in the case of the 30-minutes model for Patient B. In this case, it is desirable to select one model (model $M1$) validating at least one migraine with a high fit, instead of a model (like model $M4$) that barely reaches the 60% of fit.

Equations 2 through 4 show the expressions corresponding to the selected models in Table 2 for 20 minutes of prediction horizon for Patient A. For Patient B, the expressions of the models in bold in Table 3 for 20 minutes are those in Equations 5 through 7. Notice that most of the expressions draw an autoregressive (AR) model. These models do not depend on the inputs u_x, and only depend on the real output, yr, and/or the predicted one, yp. These models cannot be implemented in a real scenario, and must be considered as purely mathematical solutions. This means that the GE did not find anything better than the output performed by the Zero-Order Hold (ZOH) in the whole solution space. For the purpose of comparison, the average fits achieved with the selected GE models are shown with those of the ZOH in Table 5.

As Table 5 shows, the ZOH model presents higher fit values for Patient A because his/her migraines are, on average, longer than those from Patient B. Longer decay curves lead to a lower error in the ZOH models.

In view of the results, the experimental models using GE for the prediction of migraine are, so far, limited to 20 minutes of prediction. Therefore, our test experiments will be limited to this prediction horizon.

As aforementioned, this work is part of a real clinical study, and one of the goals in the envisioning of GE models is the implementation of low power consuming predictive models to be executed in the sensing motes. As stated above, this is not feasible using AR models due to their computing complexity; thus, we select the best no-autoregressive models found by the GE using the migraines that lead to the AR ones in Tables 2 and 3. The substitute models are shown in Equations 8 through 10.

The no-autoregressive solution for model $M4$ from Patient A is the one in Eq. 8. With this expression, using models $M3$, $M4$ and $M11$ for Patient A and 20 minutes of prediction horizon, only EDA (u_2) and HR (u_3) inputs are needed to predict migraines. The no-autoregressive solutions for models $M2$ and $M6$ from Patient B are the ones in Equations 9 and 10. In this case, using models $M2$, $M5$ and $M6$ for Patient B, EDA, HR, and SpO2 (u_4) are needed to predict migraines with 20 minutes in advance. None of the selected models requires the FFT functions, which contributes to generate easily implementable and low power consuming models.

Table 4: Expression for the GE models for Patient A (shadowed cells) and Patient B (white cells). The last three expressions are the no-autoregressive models used in substitution of the autoregressive ones for Patient A and B.

$y_p[k+1] = y_r[k-20] + exp\left(log\left(49 * 10^{-5} + u_2[k-108]\right) - y_r[k-113] + sin\left(u_2[k-114]\right)\right)$ (2)	
$y_p[k+1] = y_r[k-20] + \frac{exp\left(\frac{u_3[k-81]}{u_2[k-70]}\right)}{u_3[k-20]}$ (3)	$y_p[k+1] = y_r[k-20] + \frac{exp\left(\frac{u_3[k-81]}{u_2[k-70]}\right)}{u_3[k-20]}$ (4)
$y_p[k+1] = \frac{y_r[k-20] * u4[k-134]}{0.33 * u4[k-134] + y_r[k-46] + diff\left(u4[k+\tau]\Big\|_{-37}^{-20}\right) + cos\left(diff\left(u_3[k+\tau]\Big\|_{-140}^{-20}\right)\right)}$ (5)	
$y_p[k+1] = y_r[k-20] + y_r[k-21] - y_r[k-43]$ (6)	$y_p[k+1] = \frac{y_r[k-20] * y_r[k-24]}{y_r[k-39] + sin\left(y_p[k-16]\right)}$ (7)
$y_p[k+1] = y_r[k-20] + \left(65 * 10^{-4} - u_3[k-20]\right) * diff\left(u_2[k+\tau]\Big\|_{-100}^{-20}\right)$ (8)	
$y_p[k+1] = y_r[k-20] + y_r[k-29] - y_r[k-45] + log(u_3[k-53])$ (9)	$y_p[k+1] = y_r[k-20] * log\left(log\left(max\left(u_2[k+\tau]\Big\|_{-136}^{-20}\right)\right)\right)$ (10)

Table 5: Comparison of the results using GE and the ZOH predictive model.

	Patient A		Patient B	
Horizon (min)	GE	ZOH	GE	ZOH
10	86.1	69.9	88.2	69.2
20	68.7	45.7	75.7	41.5
30	52.7	28.9	68.2	18.1

3.2.1 Improvement techniques

In the following lines, the benefits of repairing the predicted signals are shown. Several repairing techniques can be applied to the output generated by the prediction models. These techniques are, in the following order: i) removal of spurious events using both time and level thresholds; ii) single event detection using a linear decider; and iii) recovering of the original desired shape of the migraine event by Gaussian fitting of the detected event. The benefits of the first and third techniques have been demonstrated in previous works [3, 5]. In this paper the single event detection has been introduced. This, when multiple events have been detected, chooses the one with a higher fit to a two semi-Gaussian curve. Figure 5 represents this situation, where in addition to the real pain curve (dark solid line), a high level event has been detected at the beginning (dotted line). This event is not considered a symptomatic curve and is removed. The fit improves from $f = -365.0\%$ to the $f = 81.1\%$ achieved with the repaired prediction (clear solid line).

As aforementioned, this improvement techniques are applied for each single prediction before the calculus of the average prediction.

3.3 Test results

In this section the selected models are applied over the test dataset. The test dataset for Patient A is composed of 5 migraines and 5 asymptomatic periods of time. For Patient B, there are also 5 migraines but 6 asymptomatic periods.

For each patient, the three models for 20 minutes are applied and then the average prediction is computed. A detection is considered as a migraine event when the fit is higher that 70% (except for Patient A, where this threshold has been downgraded to 68.5%, due to this is the average fit achieved by the selected models in training). The metric to evaluate the goodness of the models is the F-value, and the results are shown in Table 6.

Models for Patient B detect all the events in the test set, and no alarm has been produced in asymptomatic periods of time. For Patient A, one false positive has occurred due to an non-reparable

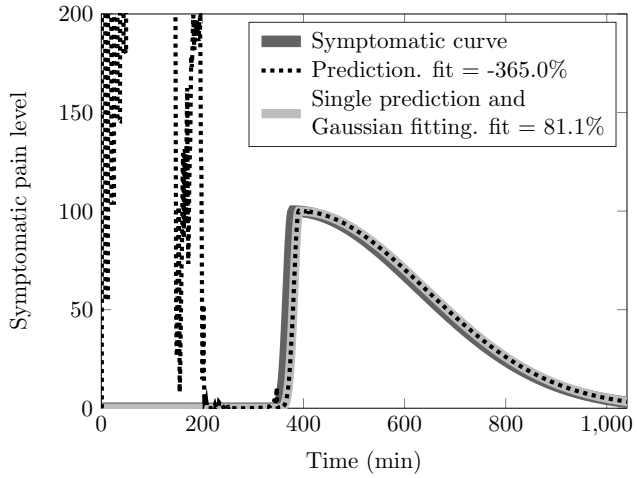

Figure 5: Example of the benefits of using improvement techniques. Here, over the original prediction, spurious and high values are removed and a single event detection is performed before applying a Gaussian fit.

Table 6: Test results for Patients A and Patient B at 20 minutes of prediction horizon at 68.5% and 70% of fit respectively.

	Patient A	Patient B
TPR (%)	60.0	100
TNR (%)	80.0	100
PPV (%)	75.0	100
F (%)	66.7	100

prediction in one of the models, and precision (PPV) falls to 75%. The sensitivity (TPR) is low due to two misclassifications. Although the F-value is not as good as expected for Patient A, test results show that the migraine prediction using GE is feasible.

Despite the maximum prediction horizon to predict migraines using GE algorithms is 20 minutes, and despite this value is far from the almost 40 minutes achieved using classic modeling techniques in previous works [3, 5], the achieved results are considered enough and sufficient to advance the drug intake and avoid the symptomatic crisis. GE algorithms are simple non-linear equations easily programmable in low power monitoring devices for real time prediction. In addition, the GE performs feature engineering and feature selection opposed to the exhaustive work

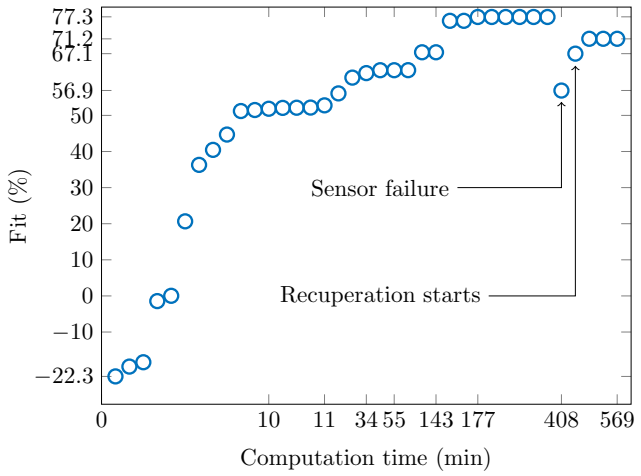

Figure 6: Simulation of the recovering capabilities of the GE system against sensor failures. After a sensor failure, the GE algorithm recovers a new model with an accuracy near to the one of the original and deprecated model.

that needs to be done in classic methods. This problem was faced by the authors in a previous work [3] and a methodology in case of sensor failure was described. Since we have the complete evolution of GE, the last population can act as a starting point in the new optimization process that happens when a sensor fails. Results obtained are excellent as shown in Figure 6.

We can consider a real Mobile Cloud Computing scenario [8] where the GE models is performed in a cloud server. The selected prediction models run in the monitoring devices, where any of the sensors can fail. In that case, the prediction model of the failing device is not valid while the sensor is out. However, upon this situation, in the remote server, the GE algorithm starts the training from the obsolete model expression avoiding the use of the immeasurable biometric variable. In this case, GE quickly finds another expression with a similar accuracy to be used while the sensor is broken. Therefore, our approach based on GE is suitable for a large-scale monitoring scenario with real-life application.

The aforementioned scenario is simulated in Figure 6, where we present the evolution of the fit in a training experiment (notice that time axis is not linear). In the 80% of the trained generations (120000 generations), one of the inputs is set to 0, and suddenly the fit falls. From that point, the GE starts looking for new expressions and in 59 minutes achieves a new one with the 86.8% of accuracy of the original (77.3%). 34 minutes later the fit is 71.2%, the 92.1% of the original. This experiment shows the robustness of the GE algorithm for migraine prediction against sensor failures in a real scenario; the GE automatically finds a new model. This is another advantage over the classic models, where the selection of a new model is an exhaustive and manual off-line work.

The utility of GE for predicting migraines has been demonstrated. Better expressions would be found training the models with a higher population. Due to computing time limits in the available cluster for this work, this was not possible and it will be tackled in the future. Also, as a future work, a penalty will be introduced in the GE training to avoid autoregressive expressions; in this way better models will be selected.

4. CONCLUSIONS

This paper demonstrates the potential of using Grammatical Evolution for migraine prediction. This work is part of a real medical study implementing ambulatory monitorization of hemodynamic variables in migraine sufferers by using WBSNs. The study shows that the migraine prediction using GE is time limited to 20 minutes in advance. This result is good enough regarding the pharmacokinetics of current treatments. Thus, an event alarm can be implemented to anticipate the intake of the drugs before the pain starts.

As opposed to classical approaches, GE computes automatically feature selection. GE obtains simple non-linear mathematical expressions for the prediction models. These expressions can be easily implemented in low power wireless monitoring devices. They can be run in real time with low computing overhead, being far faster and more efficient than implementations of classical methods. Furthermore, our results show that this alternative is robust against sensor failures, providing automatically new models that avoid the corrupted sensor.

References

[1] H. Alemdar and C. Ersoy. Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54(15):2688–2710, 2010.

[2] M. Alfonseca and S. Gil. Evolving a predator–prey ecosystem of mathematical expressions with grammatical evolution. *Complexity*, 20(3):66–83, 2015.

[3] Anonymized. Anonymized title for blind review. *Anonymized journal for blind review*, 2015.

[4] Anonymized. Anonymized source for blind review, 2016.

[5] Anonymized. Anonymized title for blind review. *Anonymized journal for blind review*, 2016.

[6] P. Arroba, J. L. Risco-Martín, M. Zapater, J. M. Moya, and J. L. Ayala. Enhancing regression models for complex systems using evolutionary techniques for feature engineering. *Journal of Grid Computing*, 13(3):409–423, 2015.

[7] R. Braojos Lopez, H. Mamaghanian, R. Vallejos, F. Javier, A. Dias Junior, G. Ansaloni, S. Murali, and D. Atienza Alonso. Ultra-low power design of wearable cardiac monitoring systems. In *Proceedings of the IEEE/ACM 2014 Design Automation Conference (DAC)*, volume 1, pages 140–145. IEEE/ACM Press, 2014.

[8] H. T. Dinh, C. Lee, D. Niyato, and P. Wang. A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless communications and mobile computing*, 13(18):1587–1611, 2013.

[9] G. Georgoulas, D. Gavrilis, I. G. Tsoulos, C. Stylios, J. Bernardes, and P. P. Groumpos. Novel approach for fetal heart rate classification introducing grammatical evolution. *Biomedical Signal Processing and Control*, 2(2):69–79, 2007.

[10] J. I. Hidalgo, J. M. Colmenar, J. L. Risco-Martin, A. Cuesta-Infante, E. Maqueda, M. Botella, and J. A. Rubio. Modeling glycemia in humans by means of grammatical evolution. *Applied Soft Computing*, 20:40–53, 2014.

[11] X. H. Hu, N. H. Raskin, R. Cowan, L. E. Markson, M. L. Berger, U. S. M. S. P. U. Group, et al. Treatment of migraine with rizatriptan: when to take the medication. *Headache: The Journal of Head and Face Pain*, 42(1):16–20, 2002.

[12] P. Kugler, D. Schuldhaus, U. Jensen, and B. Eskofier. Mobile recording system for sport applications. In *Proceedings of the 8th international symposium on computer science in sport (IACSS 2011), Liverpool*, pages 67–70, 2011.

[13] M. Linde, A. Gustavsson, L. Stovner, T. Steiner, J. Barré, Z. Katsarava, J. Lainez, C. Lampl, M. Lantéri-Minet, D. Rastenyte, et al. The cost of headache disorders in europe: the eurolight project. *European journal of neurology*, 19(5):703–711, 2012.

[14] R. B. Lipton and A. I. Scher. Epidemiology and economic impact of migraine. *Current Medical Research and Opinion®*, 17(S1):s4–12, 2001.

[15] Nonin. Nonin website. http://www.nonin.com/Home. Accessed: 2016-01-01.

[16] PLUX. PLUX-Wireless Biosignlas website. http://www.biosignalsplux.com/index.php/en/. Accessed: 2016-01-01.

[17] R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza. *A field guide to genetic programming*. Lulu. com, 2008.

[18] C. E. Rasmussen and H. Nickisch. The gaussian processes web site. http://gaussianprocess.org/gpml/code. Accessed: 2016-01-01.

[19] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[20] C. Ryan and M. O'Neill. Grammatical evolution: A steady state approach. *Late Breaking Papers, Genetic Programming*, 1998:180–185, 1998.

[21] J. C. Salinas-Hilburg, M. Zapater, J. L. Risco-Martin, J. M. Moya, and J. L. Ayala. Using grammatical evolution techniques to model the dynamic power consumption of enterprise servers. In *Complex, Intelligent, and Software Intensive Systems (CISIS), 2015 Ninth International Conference on*, pages 110–117. IEEE, 2015.

[22] L. J. Stovner and C. Andree. Prevalence of headache in europe: a review for the eurolight project. *The journal of headache and pain*, 11(4):289–299, 2010.

[23] G. Surrel, F. Rincon, S. Murali, and D. Atienza. Design of ultra-low-power smart wearable systems. In *Test Symposium (LATS), 2015 16th Latin-American*, pages 1–2. IEEE, 2015.

[24] A. Tobola, O. Korpok, H. Leutheuser, B. Schmitz, C. Hofmann, M. Struck, C. Weigand, B. Eskofier, A. Heuberger, and G. Fischer. System Design Impacts on Battery Runtime of Wearable Medical Sensors. In D. Novák, editor, *Proceedings of International Conference on Mobile and Information Technologies in Medicine and Health*, pages 1–4, 2014.

[25] A. Tobola, F. J. Streit, C. Espig, O. Korpok, C. Sauter, N. Lang, B. Schmitz, C. Hofmann, M. Struck, C. Weigand, et al. Sampling rate impact on energy consumption of biomedical signal processing systems. In *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on*, pages 1–6. IEEE, 2015.

[26] WHO. Headache disorders, fact sheet n°277. http://www.who.int/mediacentre/factsheets/fs277/en/, 2012. Accessed: 2016-01-01.