Feature Modeling for Interpretable Low Back Pain Classification Based on Surface EMG

Srhoj-Egekher, Vedran; Cifrek, Mario; Peharec, Stanislav

Source / Izvornik: **IEEE Access, 2022, 10, 73702 - 73727**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.1109/access.2022.3190102

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:184:285921

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 June 2022, accepted 4 July 2022, date of publication 12 July 2022, date of current version 18 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3190102

RESEARCH ARTICLE

Feature Modeling for Interpretable Low Back Pain Classification Based on Surface EMG

VEDRAN SRHOJ-EGEKHER[1], (Student Member, IEEE), MARIO CIFREK[1], (Senior Member, IEEE), AND STANISLAV PEHAREC[2]

[1]Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
[2]Faculty of Medicine, University of Rijeka, 51000 Rijeka, Croatia

Corresponding author: Vedran Srhoj-Egekher (vedran.srhoj-egekher@fer.hr)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the the Faculty of Electrical Engineering and Computing, University of Zagreb and performed in line with the ethical principals and recommendations of the Declaration of Helsinki (and its revisions).

ABSTRACT Low back pain (LBP) is a global health-problem phenomenon. Most patients are categorized as non-specific, thus requiring an individualized approach which still poses a major challenge. In this paper, sEMG recordings from two pairs of lumbar muscle sites were collected during an isometric trunk extension exercise. Ninety-one subjects were included in the study; 29 patients with non-specific chronic LBP (CLBP), 25 patients with radiculopathy (RLBP), and 37 control healthy subjects (HS). Six best-performing time-domain raw features were employed to model contextual secondary feature groups. Neuromuscular LBP characteristics were described with coordination, co-activation, trends, and fatigue measures. Altogether, a set of 327 secondary features was created where inputs into the classification models were further refined by employing neighborhood component analysis (NCA). NCA effectively reduced the number of features (<20 components), alongside preserving them in the original interpretable domain. A set of 23 different classifiers was employed and explored, resulting in classification accuracy of 0.94 for HS vs. LBP, 0.89 for HS vs. CLBP, 0.98 for HS vs. RLBP, and 0.89 for CLBP vs. RLBP differentiation. High median precision (0.97) and sensitivity (0.99) across all classifiers for HS vs. RLBP differentiation was obtained, with only three feature components utilized (out of 327). Support vector machines (SVM) and *k*-nearest neighbor (*k*NN) based classifiers consistently demonstrated best classification results. Different profiles of CLBP patients were presented and discussed. The suggested method demonstrated the potential for patients' subgrouping and subsequent more individualized rehabilitation treatments, backed by medical interpretations through feature modeling.

INDEX TERMS Chronic low back pain, clinical decision support systems, feature modeling, interpretability, machine learning, patients differentiation, radiculopathy, surface electromyography.

I. INTRODUCTION

Low back pain (LBP) has been clearly identified and confirmed as one of the major global public health problems for decades now, ranked as the number one cause of years lived with disability [1]. This inevitably leads to a significant number of all people experiencing some form of LBP during their lifetime, with a currently reported global yearly prevalence of over 500 million people at any time [1]. This phenomenon, taking large scale, not only affects the health and abilities of the individuals but has tremendous socio-economic implications, consequently incurring significant costs [2].

Another revealing insight is that, for about 90% of patients with LBP disorders, it can not be clearly stated the cause of the pain, thus such patients are being classified as *non-specific* LBP with treatments directed primarily on reducing pain itself and its consequences. Among patients with LBP disorders, 5-10% of patients can usually be related to radicular syndrome [3]. Such an approach, with

The associate editor coordinating the review of this manuscript and approving it for publication was Khin Wee Lai.

no specification, cannot tackle the real causes of LBP problems and more subtle musculoskeletal changes with neuromuscular adaptations in behind [4]. Equally, guarantees of long-term positive outcomes for medical treatments and rehabilitation are thus less likely to hold. This led to hypothesizing that (non-specific) LBP patients are not a homogeneous group [5], but more likey a more individualized approach in diagnostics and rehabilitation is needed to account for diversity and complexity in behind [4], [5]. An additional challenge is a high level of reported ambiguities in understanding muscle recruitment mechanisms in LBP patients (as thoroughly summarized by Dieën *et al.* [6]), thus becoming an obstacle in supporting better treatment outcomes overall. In that course, considering and introducing multiple factors in terms of characteristics describing LBP could enable identifying more specific motor control patterns, potentially leading to a more homogeneous subgrouping of LBP patients. This would be of high importance.

There has been a history of different methods proposed and techniques employed to assess, predict or discriminate LBP in patients, as summarized in several review studies [6]–[8]. Some of these methods are relying on different imaging techniques, like highly invasive X-ray or Computed Tomography (CT) [9], less invasive ultrasound imaging [10], or non-invasive, but expensive MRI [11]. Nonetheless, the most commonly used approach to assess LBP in patients is through surface electromyography (sEMG). sEMG is a well-established technique that enables non-invasive recording of muscle activity. The myoelectric activity reflects the neuromuscular adaptations in LBP patients, thus enabling more insights into motor control coordination and co-activation strategies among muscles. [6], [12]–[16].

The advancements and an ever-increasing interest in machine learning (ML) and artificial intelligence (AI) approaches, with their applicability to different domains, were also reflected in different research efforts tackling LBP problems. In the recent period (2020-2022), several review studies have been presented focusing on the applications of ML and AI in the detection, classification, treatment, or management of LBP [8], [17], [18]. Although the respective review studies examined a large number of publications and research works within the LBP domain, it was observed that the number of papers tackling the LBP by exploiting ML and AI techniques, based on sEMG only, is still relatively small. Hence, Tagliaferri *et al.* [8] reported in their review study only a few sEMG-only based research papers to be considered with a ''fair quality'', i.e. with no shortcomings [19], [20], whereas the remaining ones were stated to be with ''poor quality''. Jenssen *et al.* [17] identified two hundred and eighty-seven publications, where, after the refinement process, fifty-three papers on chronic pain research and machine learning were reviewed. Among these, only two papers were reported to deal with LBP by utilizing sEMG exclusively, whereas the additional three papers where combining sEMG with other types of information. In the most recent review study, by D'Antoni *et al.* [18], the AI in the context of computer-aided diagnosis (CAD) in CLBP was reviewed. Similarly, only two studies were reported to deal with LBP by employing sEMG only [21], [22], and one study combined sEMG with kinematic variables [23]. General conclusions drawn from these studies are suggesting that ML/AI in LBP is in its infancy [8] and that more attention is needed on ML/AI approaches to the LBP prediction, subgrouping, and treatments with prospects for the positive rehabilitation outcomes [8], [17].

Available cross-sectional studies tackling LBP detection or classification by means of sEMG only, were mostly focusing on separation between non-specific chronic LBP (CLBP) and healthy subject (HS) groups [12], [20], [21], [24]–[27]. Nonetheless, for the presented papers at hand, the classification accuracy for separating CLBP from HS was ranging between 80% in [21] to 98.04% in [20] (for one of the feature subsets). In a number of studies, sEMG is combined with kinematic tracking in order to provide more controlled insights into the related muscle activation and movement patterns [16]. This enables expansion of the set of characteristics (beyond sEMG features only) that are subsequently employed for detection or discrimination of LBP patients [23], [28], [29]. Generally, it is observed that such studies, combining sEMG and kinematic parameters, were resulting in better classification accuracy results compared to sEMG-only ones. These studies were also, among the rare ones, to tackle further discrimination within CLBP groups. Dankaerts *et al.* [28] reported overall classification success in discriminating healthy controls and two clinical subgroups of patients (Flexion Pattern, FP, and Active Extension Pattern, AEP) with 96.4% accuracy, however, the accuracy of 84.1%, if only sEMG features were utilized. Olugbade *et al.* employed classifiers in recognizing different LBP levels (low and high pain levels) in patients, with the accuracy of 94% for the best-performing support vector machines (SVM) classification model tested (random forests, RF, were also employed). Another study, moreover utilizing sEMG features only, employed decision trees in differentiating CLBP and patients with radiculopathy (RLBP) from the HS group, with at best accuracy of 86.8% [30].

Furthermore, some of the studies tackled the prediction of rehabilitation outcomes. Liew *et al.* [23] employed three different binary classification models for differentiating between patients with current LBP and with LBP in remission, alongside a third control group, where the best-reported area under the receiver operator curve (AUC) of 96.7% was obtained for differentiation between two LBP groups (current and remission). A similar cohort study was presented by Jiang *et al.* for discriminating the ''responding'' against the ''non-responding'' LBP patients' treatment outcomes [19], by utilizing sEMG-only features, with best-reported accuracy of 96.67%.

Dominant classification models, in presented studies, were based on discriminant analysis [12], [24]–[27], where in more recent works, SVM classifiers are gaining more spotlight [19], [20], [29]. There have been also some

attempts to apply deep learning methods in sEMG-based LBP recognition with convolutional neural networks (CNN) [22], reporting a high average classification accuracy of 92.9%, but still without such approaches gaining more traction, especially if only sEMG features were considered. Such an example of employing non-sEMG parameters was a study of recognizing LBP individuals from healthy ones with kinematic parameters as input into the deep learning neural network with long short-term memory (LSTM), and reporting both classification precision and recall up to 97.2% [31]. Further, some papers were investigating deep learning methods in conjunction with sEMG, where LBP was not directly tackled, but the target was in real-time forecasting of sEMG features for trunk muscle fatigue [32]. Additional examples of studies employing different neural network techniques, as part of CAD in CLBP, can be found in [18], where the presented studies were dominantly based on MRI and other medical imaging techniques.

It is also worth noticing that in most cases studies, relying on the dynamic functional type of exercise tasks, were the ones also demonstrating the best classification results [19], [20], [23], [26], [29]. Where also, in many cases, these studies were collecting inputs from multiple muscles or muscle sites [23], [26], [28]. Some of the examples, where good classification results were obtained with a fewer number of sEMG channels, were presented in [20], [29], where furthermore, a simple isometric Biering-Sørensen exercise [33] was employed in [22].

Research in this paper has been motivated by still an obvious lack of success in defining subgroups of LBP patients, consequently leading to the inability to define effective treatments for non-specific LBP patients [28]. Moreover, as stated in [34], it is considered the ''Holy Grail'' of low back pain research to establish such methods that would enable subgrouping of LBP patients, backed with a consistent medical interpretation. By examining the so far reported results and state-of-the-art, we are hypothesizing that an additional contribution to successful subgrouping of LBP patients can be supported by introducing more comprehensive multifactorial LBP classification models, thus also capturing more individually-biased expressions of LBP occurrences and avoiding a common pitfall of ''one-size fits all'' approach [5], [34]. A similar view was shared in a recent work by Fallaa *et al.* [35] pointing to possibilities of ML to harness the variability of patients' presentations to enhance clinical predictions, as well as highlighting the importance of not relying on single features when characterizing patients, given the variability of physiological adaptations present in people with spinal pain. Moreover, current procedures leading to subgrouping of LBP patients did not lead to successfully tailored treatment approaches [36], thus, still waiting for more meaningful and interpretable systems able to capture the variability of physiological and neuromotor adaptations in LBP patients. Subsequently, another key ingredient in our proposed method is an attempt to reflect the LBP specificities directly within classification features themselves, more precisely, to model such features that (1) contain contextual information pertained to LBP, and (2) are interpretable by medical experts. Namely, currently, it is still a predominant case that features inputs into the classifiers are simplistic features or measures (e.g. straightforward RMS, MDF or rectified EMG values as utilized in [20], [23], [25], [29], [30]), meaning, not specific to LBP phenomenon itself and its complex coordination, co-activation and compensation mechanisms in any directly modeled manner. Thus, our overall hypothesis is that LBP complexity requires a more elaborated classification modeling approach to enable insightful inference. This idea follows the concerns and insights established throughout the review studies by Dieën *et al.* [6]), Geisser *et al.* [14], and pace-setting works by Hodges [4], [37]. In this research, the complexity focus is put on feature modeling, not on the complexity of the acquisition systems or exercise protocols.

In order to validate our hypothesis and the proposed method construction, we have opted for the following setup: sEMG features only (thus avoiding bias from other non-myoelectric values), simple isometric trunk extension exercise (thus avoiding bias from dynamic functional tasks that have demonstrated certain LBP discriminating power), only four sEMG channels for myoelectric activity acquisition (thus avoiding contribution to discrimination success usually pertained to multi-electrode setups), classifiers employed *as-is* out-of-the-box (thus avoiding bias from additional classifiers optimization or fine-tuning). The rationale behind a such proposed setup is as follows: if this minimum setup can demonstrate success in detecting LBP patients and provide insights into LBP groups at hand (namely, CLBP and RLBP) with satisfactory differentiation accuracy, alongside preserving the possibility to interpret the results, then it is reasonable to assume that further improvements and optimization can contribute even further to the overall goal of enabling effective and meaningful subgrouping of LBP patients.

Finally, we consider this work as a step forward in building clinical decision support systems (CDSS) that may be used by medical professionals in clinical practice as part of more concrete medical guided-based recommendation systems, with strong interpretable models. Moreover, such explainable models are mandated by European Union's General Data Protection Regulation (GDPR), as pointed out in work by Jović *et al.* [38].

II. MATERIALS AND METHODS

A. SUBJECTS

Three different cohorts of male volunteers were involved in our research based on the presence of low back pain or related pathologies. All subjects were examined individually by the medical professional and, following up their medical history track and diagnostic check, each subject was assigned to one of the three groups: HS, non-specific CLBP, and RLBP. Inclusion criteria for CLBP were defined as daily or almost daily pain that lasted at least three months prior to

TABLE 1. Number of subjects (*N*) for HS, CLBP, RLBP cohorts with demographic (age) and biometric (height, weight, body mass index - BMI) details, together with Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RDQ) outcome and Visual Analogue Scale (VAS) pain scoring, are shown as the mean value and standard deviation (mean ± std).

	HS (N = 37)	CLBP (N = 29)	RLBP (N = 25)
Age (years)	32.62 ± 9.17	34.1 ± 10.12	36.4 ± 9.0
Height (cm)	180.76 ± 5.18	184.03 ± 7.63	182.56 ± 5.52
Weight (kg)	82.84 ± 8.42	87.24 ± 10.39	86 ± 9.21
BMI (kg/m^2)	25.35 ± 2.23	25.70 ± 2.0	25.81 ± 2.64
ODI (0-100)	7.14 ± 8.1	15.59 ± 9.49	33.04 ± 14.1
RDQ (0-24)	1.86 ± 2.57	5.34 ± 3.91	11.24 ± 4.76
VAS (0-10)	1.65 ± 2.02	2.74 ± 1.84	5.6 ± 2.14

measurements [16], [28], [39]. For RLBP patients, the main inclusion criteria were clinical symptoms of radiculopathy with a positive Lasegue sign on one side lasting at least fourteen days [26], [40], [41]. The overall exclusion criteria for subjects were spinal deformation, spinal surgery, spondylolisthesis, spinal stenosis, or spinal injuries, altogether with the indication that there were no accompanying systemic diseases [16], [26]. Following the defined procedure, ninety-one male subjects were detected as suitable for measurement protocol and sEMG data acquisition. Alongside collecting the biometric and demographic details, subjects were provided intake of standard questionnaires, namely, Oswestry Low Back Pain Disability Questionnaire (ODQ), Roland-Morris Disability Questionnaire (RDQ), and Visual Analogue Scale (VAS) for pain scoring, to assess their own LBP related functional disability and pain perception status [13], [16], [28], [42]. All these metadata are summarized in Table 1. The data has been collected at the Biomechanical laboratory of Polyclinic for Physical Therapy and Rehabilitation, Pula, Croatia. The collection time was in the morning from 9:00 am to 12:00 pm.

The whole experiment was approved by the Ethics Committee of the University of Zagreb, Faculty of Electrical Engineering and Computing, and informed consent was received from each subject.

B. MEASUREMENT PROTOCOL AND DATA ACQUISITION

Measurement protocol was based on the isometric trunk extension exercise in the configuration of Roman chair (a variant of *Biering-Sørensen* test) [30], [33], [42], as shown in Fig. 1. Myoelectric muscle activity was recorded with FREEEMG (BTS, Milano, Italy) system with wireless surface EMG probes enabling free movement of subjects during the measurement. Each surface EMG probe applied had a pair of pre-gelled Ag-AgCl 10 mm diameter electrodes (Ambu-Blue, Sensor, and Ballerup, Denmark). Prior to applying the electrodes onto the skin, the respective surface was prepared in accordance with SENIAM recommendations with surface cleansing and electrode-skin impedance controlled (<5 kΩ) [43]. The main characteristics of the EMG recording system were given by: differential amplifying with bandpass filtering (20-400 Hz), differential input impedance

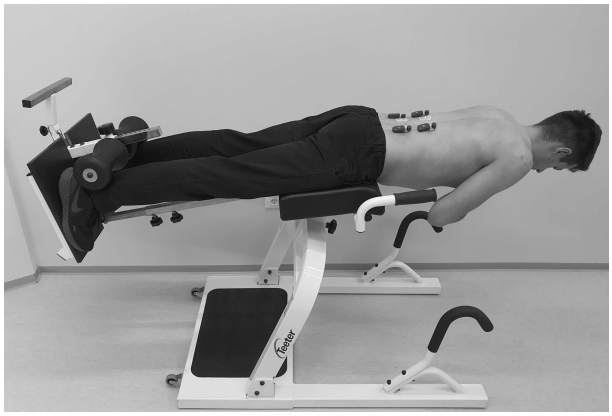

FIGURE 1. Tilting device in the Roman chair configuration used for testing with the subject placed in the trunk extension position.

>100 MΩ, CMRR >100 dB (at 65 Hz) with 1000 Hz sampling frequency using the 12-bit A/D converter. Outside of the 20-400 Hz frequency range, the contribution to the sEMG spectrum is deemed negligible [44].

To commence a testing procedure, each subject was familiarized with the procedure, a tilting device, and instrumentation. Further, each subject was asked to stand upright until electrodes were positioned. Four sEMG channels were acquired in the paraspinal lumbar region, placed over the upper lumbar erector spinae (ULES) and lower lumbar erector spinae (LLES) muscle sites, bilaterally (left and right) in respect to the vertical spine axis (Fig. 2). While standing in the upright position without footwear, the distance between the floor and *anterior superior iliac spine* was measured. Depending on the measured distance, the standing pad of the tilting device was adjusted so that the toes, back of the lower leg (above the Achilles tendon), and pelvis (together with the upper thigh) became the only body parts in contact with the tilting device and thus creating the supporting points. The subject was instructed to stand on the tilting device and to hold hands crossed having palms placed on the chest. Upon the subject's verbal confirmation, the medical staff gradually tilted the device until a horizontal position was reached. To ensure static contractions of lower back muscles each subject was asked to maintain the horizontal position as stable as possible. The weight of the subject's upper body was used to induce muscle fatigue. No differences in measurement and exercise protocol, concerning the subjects' medical status, were demonstrated and all subjects were equally asked to sustain the isometric contraction as long as possible or until a maximum exercise duration of 180 seconds was reached. When the exercise end was reached, medical staff returned the subject to an upright position [30].

C. METHODS PROCEDURE OVERVIEW

Our research and proposed method consists of several steps that are presented in more details in the sections III−VI. Here, we shortly list the main steps involved:

1) Extracting raw (primary) features (section III)

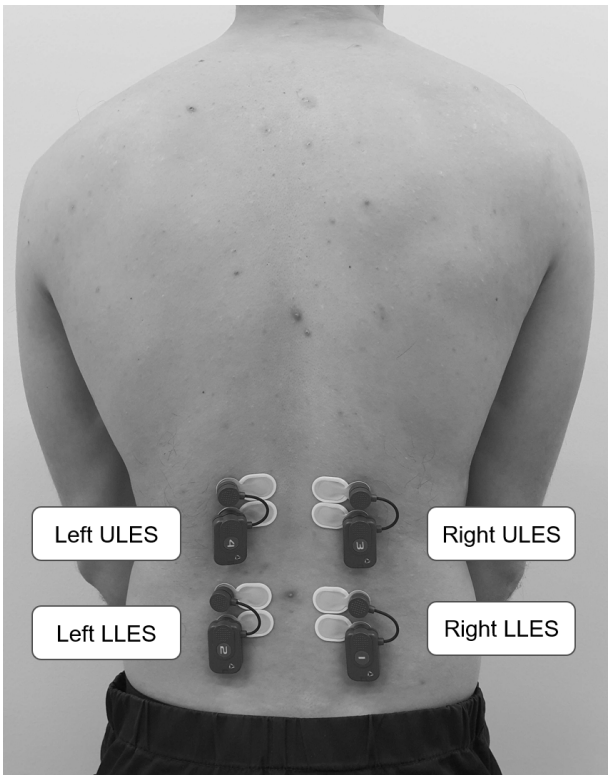

FIGURE 2. Placement of two bilateral pairs of wireless electrodes on the L1-L2 (ULES) and L4-L5 (LLES) positions. Altogether four channels.

2) Constructing contextual (secondary) feature groups (section IV)
3) Feature analysis and selection (section V)
4) Classification and validation procedure (section VI)

Steps 3) and 4) were repeated iteratively in order to select and validate feature vector sets and propose the final model.

III. RAW FEATURES

A. SEMG SIGNAL PROPERTIES AND LIMITATIONS

Surface myoelectric signals are generally characterized by non-stationary [45]–[47] and non-linear [48]–[50] stochastic properties as a result of not having satisfactory models describing a deterministic nature of complex neural control for muscle contractions [51]. Thus, different segments within sEMG time sequence have different statistical properties and carry different pieces of information relevant for respective detection and inference of LBP conditions. This can be well noticed in multi-electrode sEMG topography measurements with dynamic exercise tasks [52]. Thus, it would be plausible to analyze consecutive portions of sEMG signals to infer certain localized time-dependent properties that would otherwise be lost. Alongside, within shorter time chunks of sEMG signal, statistical properties are changing much slower and wide-sense stationarity (WSS) can be assumed. This primarily holds for myoelectric signals derived from isometric contractions with duration of $500 - 2000$ ms [45], [53], [54], whereas for dynamic type of contractions these assumptions may be violated [54]–[56].

Based on these insights, in this study time chunks of $L = 1000$ ms were applied to calculate primary features with overlapping step $s = 50$ ms (the equivalent of 95% overlapping between neighboring time chunks).

B. PRIMARY FEATURES EXTRACTION

As the first step in our proposed procedure, simple raw features (*primary features* further in the text) were calculated directly from the raw sEMG signals. These primary features in subsequent steps were used to create more LBP-meaningful secondary features. Depending on the domain where features are calculated, they are generally categorized as time-domain (TD), frequency domain (FD), time-frequency domain (TFD), and spatial domain (SD) based features. In this study, we have opted mostly for the time domain-based features due to their simplicity and speed of calculation. The full list of primary features being utilized, with the originating domain and types of information carried by the respective features, is given in Table 2. We have used thirteen TD features: zero crossing (ZC), signal slope change (SSC), Willison amplitude (WAMP), mean absolute value (MAV), integrated EMG (IEMG), variance (VAR), root mean square (RMS), waveform length (WL), log detector (LD), kurtosis (KURT), skewness (SKEW), permutation entropy (PE) and relative variance difference (RVD). The only FD feature used is the power spectrum median frequency (MDF). This feature list, among others, contains all features presented by Hudgin [57] or Du [58]. Although respective feature vectors (namely, Hudgin's and Du's) are primarily used in real-time applications exploiting sEMG, like human-machine interfaces (HMI), it was worth also exploring and comparing the applicability of similar feature sets in the domain of LBP detection and patients' differentiation.

Detailed overviews of these and other common features created in TD, FD, TFD, and SD can be found in [59]–[62]. The presented features, irrespective of being created in time or (time-)frequency domain, can reflect certain properties that are natively expressed in another domain. For instance, some of the TD features, like ZC or SSC are closely related to the frequency properties of the respective signal, that are originally observed in FD. Also, as part of this study, we introduced an additionally constructed RVD feature which is presented in more detail in section III-B2, together with the less commonly used PE feature (section III-B1).

Our primary features are calculated on a sequence of raw sEMG time chunks ($L = 1000$ ms, $s = 50$ ms) where for each feature (Table 2) the output is a new sequence, effectively sampled due to step s used. This new feature sequence then holds more pronounced respective domain (TD or FD) information, by reflecting the progression of myoelectric properties that are expressed in terms of the specific primary feature used.

TABLE 2. Primary features used in this study with feature name, feature domain (TD, FD) and type of information carried (frequency, energy, activation, complexity, higher order statistics - HOS).

ID	Domain	Feature name	Type of information
1	TD	Zero crossing (ZC)	Frequency information
2	TD	Signal slope change (SSC)	Frequency information
3	TD	Willison amplitude (WAMP)	Frequency information
4	TD	Mean absolute value (MAV)	Energy and complexity
5	TD	Integrated EMG (IEMG)	Energy and complexity
6	TD	Variance (VAR)	Energy and complexity
7	TD	Root mean square (RMS)	Energy and complexity
8	TD	Waveform length (WL)	Energy and complexity
9	TD	Log detector (LD)	Energy and complexity
10	TD	Kurtosis (KURT)	HOS (non-linearity)
11	TD	Skewness (SKEW)	HOS (non-linearity)
12	TD	Permutation entropy (PE)	Non-linear complexity
13	FD	Power spectrum median frequency (MDF)[a]	Frequency information
14	TD	Relative variance difference (RVD)[b]	Activation and complexity

[a]MDF is the only FD feature used in this study.
[b]RVD is a newly introduced feature in this study.

1) PERMUTATION ENTROPY (PE)

Entropy is a non-linear measure of signal complexity. There are different formulations and parameters that can be used to estimate the entropy for the given signals. We have chosen permutation entropy - PE [63]. With PE the relative occurrences of different patterns or *motifs* can be quantified. The strong side characterizing PE is the ability to tackle non-linear properties of sEMG signals together with demonstrating robustness in eliminating potential negative effects coming from interferences. The main downside is the computational complexity with calculation taking roughly two orders of magnitude more time compared to other TD features [63], [64].

The core of PE is in choosing n consecutive points of samples and constructing an n-dimensional vector. The obtained signals are then sorted in ascending order. The resulting number of permutations in the newly created sequence is of magnitude $n!$. The probability statistics for the permutations and combinations over the entire time series is calculated [64]. The probability is represented as $p(\pi)$, where π stands for different permutations. The expression is given with the following,

$$PE = -\sum_{\pi=1}^{n!} p(\pi) \ln(p(\pi)). \tag{1}$$

The given formulation and computational implementation is following the original work done by Bandt and Pompe [65]. In order to calculate PE, it is necessary to select n (n-th order of permutation entropy) where the higher order generally enables better estimation of the complexity of the underlying dynamic system. Though, on finite-size signals, too high order can result in underestimating the system complexity due to missing some of the ordered samples on a given time chunk. It is shown that an empirical formula $5 \times (n + 1)! < L$ for calculating order n provides good results [66]. Thus, in our case, with the time chunk length of $L = 1000$ ms, $n = 4$ was chosen.

2) RELATIVE VARIANCE DIFFERENCE (RVD)

Relative variance difference is an additional feature constructed as part of our research and introduced in this paper. It is constructed with a goal to enable detecting and emphasizing the relative changes in localized sEMG signal energy (related also to amplitude changes) and specific to localized contraction properties. The relative measure is introduced having in mind the character of myoelectric manifestations recorded on the surface of the skin where the absolute values of sEMG signals may quite differ not only among different subjects but also among different channels recorded from the same subject. The presented measure (feature) is given by

$$\begin{aligned} RVD(x_k) &= \frac{Var(x_{k+}) - Var(x_{k-})}{Var} \\ &= \frac{\frac{1}{W-1}\left(\sum_{i=k}^{k+W} x_i^2 - \sum_{i=k-W}^{k} x_i^2\right)}{\frac{1}{N-1}\sum_{i=1}^{N} x_i^2} \end{aligned} \tag{2}$$

where for k-th step in the time sequence, a variance difference around time sequence point x_k is calculated for the prior ($Var(x_{k-})$) and posterior ($Var(x_{k+})$) time segment of length W. In this study, segment length $W = 1000$ ms was used. Variance difference is normalized against the overall sEMG signal (channel) variance.

IV. CONTEXTUAL FEATURES

A. SECONDARY FEATURES CONSTRUCTION

The rationale behind introducing additional feature construction, on top of the primary features, was the intention to provide functional and descriptive features that reflect the concrete LBP domain-related challenges with the ability to interpret the results. One of the main tasks was to enable the detection of LBP patients and consequently provide more specific differentiation of patients within LBP groups as, nowadays, this poses one of the main challenges in dealing

with LBP patients and rehabilitation treatments [5], [6], [28], [34]. To achieve this, primary features were used to construct contextual features (*secondary features* further in the text) that are organized into the following feature groups:

1) Coordination measures
2) Co-activation triggers
3) Trends
4) Fatigue-related indices

B. COORDINATION MEASURES

Coordination-related features were introduced and proposed to detect a relationship among the channels in different muscle locations. Previous studies reported differences or asymmetry in myoelectric patterns between the left and right for patients with LBP compared to healthy subjects [24], [52], [67]. The main coordination relationship investigated in our research was the bilateral coordination, i.e. time-related dependency between the left and right side of paraspinal muscles (ULES and LLES) in the lumbar region (Fig. 2). In that course, relationships for the left-right ULES, as well as left-right LLES, were tracked separately, resulting in two features per subject per single primary feature employed.

The aforementioned bilateral muscle relationships were established by employing different distance metrics to measure how closely each left side time sequence is in alignment with the same corresponding sequence (i.e. derived from the same primary feature) on the right side. This way, a macroscopic muscle coordination indicator of myoelectric similarity between muscle locations (channels) was defined. Calculations were conducted only between the muscle location pairs for the same subject, thus sequences were of the same length inherently. Following distance metrics were employed:

1) Euclidean distance
2) Correlation distance
3) Dynamic time warping (DTW)
4) Spearman distance
5) Mutual information (MI)

Thus, overall ten secondary features per single primary feature were constructed, $N_{Co} = 10$.

1) EUCLIDEAN DISTANCE

Euclidean distance is a classic, very common distance measure between two points in the vector space. In our case, distance measure was directly employed onto two, same length, primary features time sequences, as in

$$d_{Eucl}(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{3}$$

where p and q stand for the respective feature sequences.

To enable meaningful comparison among subjects, the impact of different time sequence lengths should be eliminated by introducing normalization or correction factors. This is achieved with the following normalization expression:

$$\hat{d}_{Eucl}(p, q) = \sqrt{\frac{\sum_{i=1}^{n} (p_i - q_i)^2}{n}} \tag{4}$$

2) CORRELATION DISTANCE

Correlation distance is calculated according to the formulation

$$d_{Corr}(p, q) = 1 - \rho(p, q). \tag{5}$$

Expression in (5) is derived from the Pearson correlation coefficient given by:

$$\rho(p, q) = \frac{\sum_{i=1}^{n} (p_i - \bar{p})(q_i - \bar{q})}{\left\{\sum_{i=1}^{n} (p_i - \bar{p})^2 (q_i - \bar{q})^2\right\}^{1/2}} \tag{6}$$

where p and q stand for the respective primary feature sequences. The correlation distance measure as such does not require any additional normalization for comparison across subjects.

3) DYNAMIC TIME WARPING

DTW is a technique that enables non-linear mapping of one signal onto another one by finding the optimal path W_0 that minimizes the warping cost [68], as in

$$d_{DTW}(p, q) = \min_{w} \left\{ \sqrt{\sum_{k=1}^{K} w_k} \Big/ K \right\}. \tag{7}$$

A warping path W is a contiguous set of matrix elements ($w_k = (i, j)_k$) that characterizes a mapping between our two primary feature sequences p and q, resulting with the minimal root sum of Euclidean distances between individual sequence points, namely $d(p_i, q_i) = (p_i - q_i)^2$. Alongside the warping path being calculated, the output of this procedure is also the similarity measure between the two sequences $d_{DTW}(p, q)$. The problem of finding the respective optimal warping path is solved by utilizing dynamic programming techniques.

A trivial example of DTW distance measure and similarity detection can be illustrated with the following two sequences $p = [0, 0, 5, 0, 0, 0, 0, 0]$, $q = [0, 0, 0, 0, 0, 5, 0, 0]$, where, for DTW, a measured distance between two sequences is $d_{DTW}(p, q) = 0$, thus perfect alignment. In case of Euclidean distance, it results with $d_{Eucl}(p, q) = 7.07$. In a core, this demonstrates the basic difference and advantage of DTW compared to a straightforward Euclidean metric, but at the cost of higher computational complexity two to three times orders of magnitude. This especially can become an important constraint when dealing with lengthy sequences [69]. However, nowadays, more powerful computational systems and advancements in DTW-based algorithms have opened new perspectives for employing this powerful technique in many other fields and applications [69].

4) SPEARMAN DISTANCE

Spearman distance is calculated as Spearman correlation which corresponds to Pearson correlation between the rank values of those two variables (in our case two feature sequences) [70]. Pearson correlation primarily deals with linear relationships successfully, whereas Spearman correlation assesses the monotonic relationships (which can be non-linear). That said, Spearman correlation results in maximum values (+1 or −1) for cases when sequences with no repeating values are acting as a perfect monotone function against each other [71].

Spearman distance uses rank values for calculation (i.e. values that correspond to the relative position of individual point within the sequence after sorting):

$$
\begin{aligned}
d_{Spear}\,(p,q) &= 1 - \frac{\sum_{i=1}^{n}\left(r_{pi}-\bar{r}_p\right)\left(r_{qi}-\bar{r}_q\right)}{\left\{\sum_{i=1}^{n}\left(r_{pi}-\bar{r}_p\right)^2\left(r_{qi}-\bar{r}_q\right)^2\right\}^{1/2}}, \\
\bar{r}_p &= \frac{1}{n}\sum_i r_{pi} = \frac{(n+1)}{2} \\
\bar{r}_q &= \frac{1}{n}\sum_i r_{qi} = \frac{(n+1)}{2}
\end{aligned} \tag{8}
$$

where r_{pi} and r_{qi} stand for rank value of sequence p and q respectively at location (point) i.

5) MUTUAL INFORMATION

MI, as a metric based on the random variable entropy measure, was also employed to quantify the relationship between two sequences. MI has some interesting properties where the ability to tackle not only linear but also non-linear properties (unlike correlation coefficient or Euclidean distance) is especially useful [72]. A distance measure derived from MI is given with following:

$$
\begin{aligned}
d_{MI}\,(p,q) &= 1 - \frac{I\,(p;q)}{H\,(p,q)} \\
I\,(p;q) &= H\,(p) + H\,(q) - H\,(p,q) \\
&= H\,(p,q) - H\,(p|q) - H\,(q|p)
\end{aligned} \tag{9}
$$

where $I(p;q)$ stands for information gain and $H(p,q)$ stands for joint entropy. Further, $H(p)$ and $H(q)$ can be recognized as marginal entropies of variable (sequence) p and q respectively, whereas $H(p|q)$ and $H(q|p)$ are respective conditional entropies. As it can be observed, MI is closely related to the information gain that is commonly used as one of the criteria in decision tree classification models [73].

C. CO-ACTIVATION TRIGGERS

Co-activation triggers, as a measure or score, were introduced in this study to detect underlying, more subtle changes and relationships that occur among muscle locations in time-localized scope and were counted throughout exercise duration. In this way, co-activation patterns in patients with LBP are quantified and explored. For each channel, two scores were calculated:

1) Co-activation time-alignments of myoelectric activity across all sEMG channels
2) Co-activation misalignments to measure a lack of such time-alignments across all sEMG channels

The procedure was set as follows:

1) Find the prominent localized peaks, for the primary feature being checked, for each channel ch_i, with $i = 1$ to 4.
2) Select one channel and iterate across all maxima (peaks) for that channel. Then, for each iterated peak, with time location loc_t, check within the neighborhood ± 250 ms if such a prominent peak is also found in other channels. For each channel where such peak is found, increase the channel counter by one in the triggers activation matrix A, $A(ch_i, loc_t) = A(ch_i, loc_t) + 1$.
3) Repeat the procedure for all channels.
4) Co-activation alignment score, for each channel i, is defined as the number of time locations (loc_t) within the triggers activation matrix for which $A(ch_i, loc_t) \geq 3$. Co-activation misalignment score is similarly defined, but for case where $A(ch_i, loc_t) = 1$, meaning, that respective time instance of trigger occurred only within that single channel.
5) Normalize resulting co-activation alignment and misalignment scores against exhibited exercise duration for the respective subject

Altogether, eight secondary features per single primary feature were constructed, $N_{Ca} = 8$.

D. TRENDS

This feature group tends to establish relationships between the start segment (first ten seconds) and the end segment (last 10 seconds) of the static contraction sEMG recordings. Trends are calculated across all primary features separately. Relations are set both within muscles themselves (start vs. end) as well as among muscle location pairs (bilateral and ipsilateral). The rationale behind this approach is at least two-fold: to account for evident intra-channel myoelectric changes as exercise time progresses, and, to account for inter-channel relationships as part of neuromuscular dependencies among muscles (i.e. muscle locations as in our case). Trend features, as part of the secondary features group, are defined as follows:

1) Relative start-end difference for left and right LLES, separately (10). Two features are calculated as max and min out of these two metrics for LLES ($N_{Tr_1} = 2$).
2) Relative start-end difference for left and right ULES, separately (11). Two features are calculated as max and min out of these two metrics for ULES ($N_{Tr_2} = 2$).
3) Ipsilateral ratio as a relative difference between LLES and ULES end and start segments, for left and right

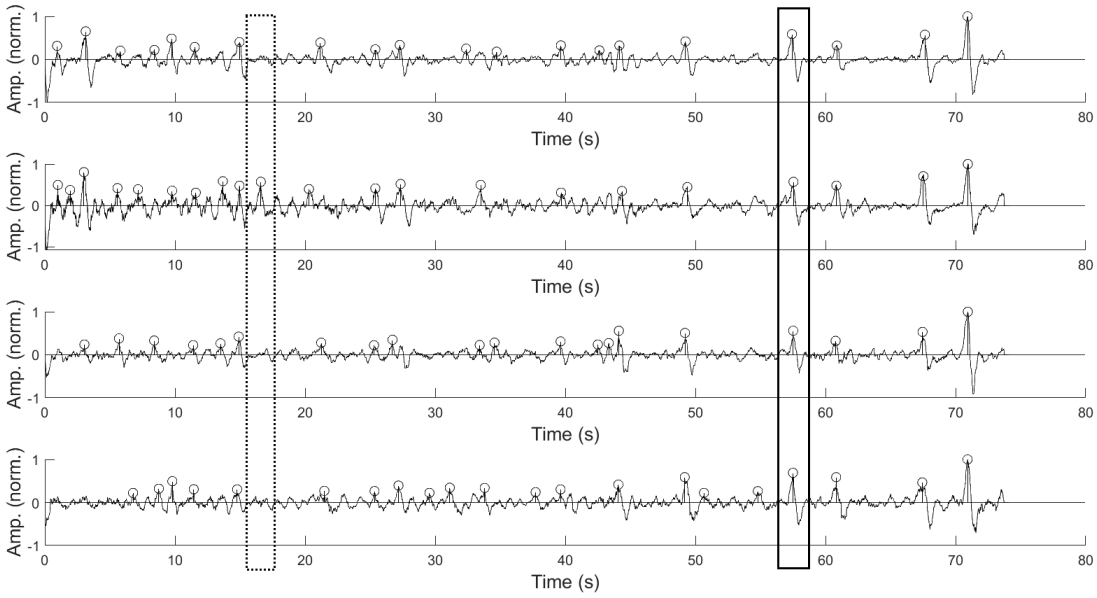

FIGURE 3. Co-activation triggers shown for all four channels, left ULES, left LLES, right ULES, and right LLES, from top to bottom respectively, taken from one healthy subject. The left vertical box, outlined with a dotted line, captures one of the occurrences with missing co-activation alignment (i.e. trigger present only on left LLES muscle location) in ±250 ms neighborhood. The right vertical box, outlined with a solid line, captures one of the occurrences with co-activation alignment (i.e. triggers found across all muscle locations). Co-activation triggers were captured via RVD (primary feature) prominent peaks, normalized to max peak value for each channel separately.

side separately (12). Two features calculated as max and min out of these two metrics for left and right ($N_{Tr_3} = 2$)

4) Absolute relative left-right difference calculated for LLES start and end, separately, and normalized with a sum of left and right (13). Thus, two relative features per one subject, ($N_{Tr_4} = 2$).
5) Absolute relative left-right difference calculated for ULES start and end, separately, and normalized with a sum of left and right (14). Thus, two relative features per one subject, ($N_{Tr_5} = 2$).

The expressions are given with:

$$Tr_1(a) = \frac{LLES_{start}(a) - LLES_{end}(a)}{LLES_{start}(a)} \tag{10}$$

$$Tr_2(a) = \frac{ULES_{start}(a) - ULES_{end}(a)}{ULES_{start}(a)} \tag{11}$$

where a takes either left or right of the lumbar region. $LLES_{start}$ or $ULES_{start}$ stand for an averaged value of the corresponding (primary) feature for the first 10 s of contraction, whereas $LLES_{end}$ or $ULES_{end}$ stand for the averaged value of the corresponding (primary) feature for the last 10 s of contraction.

$$Tr_3(a) = \frac{LLES_{end}(a) - ULES_{end}(a)}{LLES_{start}(a) - ULES_{start}(a)} \tag{12}$$

where a takes either left or right side of LLES and ULES region, separately, for start or end segments of the contraction.

$$Tr_4(seg) = abs\left(\frac{LLES_L(seg) - LLES_R(seg)}{LLES_L(seg) + LLES_R(seg)}\right) \tag{13}$$

$$Tr_5(seg) = abs\left(\frac{ULES_L(seg) - ULES_R(seg)}{ULES_L(seg) + ULES_R(seg)}\right) \tag{14}$$

where seg stands for either start or end segment of the respective sEMG recording, separately, whereas L and R stand for the left or right side of these corresponding ULES or LLES regions, respectively.

Altogether, ten secondary features per primary feature were created, $N_{Tr} = 10$, and normalized with respect to the exhibited exercise duration.

A visual inspection of time-evolving changes for the underlying primary features, based on which trends are calculated, are shown in Fig. 4. For most of the trends, a time-dependent decrease is observed. For SKEW, PE and RVD, decreasing trend behavior is not explicit or observed in this example.

E. FATIGUE INDICES

Muscle fatigue analysis by means of sEMG has a wide and well-accepted usage [74]. Moreover, muscle fatigue assessment, by exploiting MDF, has been also widely used as an indication for different LBP conditions in patients [42], [75], [76]. In that context, we had included in our study a set of extended fatigue indices, going beyond previously presented trend features (section IV-D) that also exploit MDF as one of the underlying primary features. Thus, first, the linear regression against MDF time changes was calculated in terms of least mean square error (LMSE) (Fig. 5). Then, indices from linear regression slope k_{sl} and initial frequency

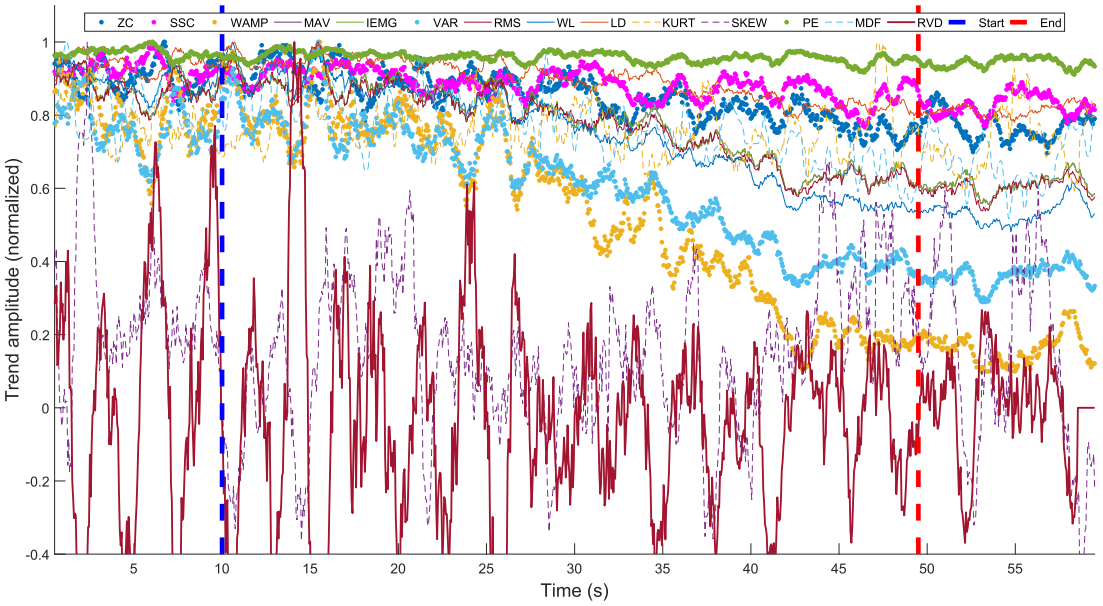

FIGURE 4. Time changes for each of fourteen primary features, calculated from the right LLES muscle location of one CLBP subject. Each primary feature trend is normalized to its max value. Vertical dashed lines represent limits for the start segment (blue), for the first 10 s, and last (red) 10 s of sEMG exercise duration, respectively. Trends for PE (green), SSC (magenta), ZC (blue), VAR (light blue), WAMP (ocher), and RVD (dark red) from top to bottom, respectively, are shown in a thicker line representation.

f_0 were derived (15):

$$y_r(x_{MDF}) = k_{sl} \times x_{MDF} + f_0. \tag{15}$$

The first subset of MDF fatigue-based features were straightforward linear regression values k_{sl} and f_0, thus eight common features were directly extracted, $N_{FI_0} = 8$ (two features per each sEMG channel).

Considering different absolute values for MDF, not only among different subjects but also among sEMG channels for the same subject (especially valid for the derived parameter f_0), additional fatigue indices were composed to reflect the relative nature between k_{sl} and f_0 parameters as fatigue is progressing. These relative indices were defined as follows:

1) Relative left-right side difference for k_{sl} for ULES and LLES, respectively, normalized with the higher of k_{sl} value between left and right. Thus, two relative features ($N_{FI_1} = 2$).
2) Relative left-right side difference for f_0 for ULES and LLES, respectively, normalized with the higher of f_0 value between left and right. Thus, two relative features ($N_{FI_2} = 2$).
3) Relative left-right side difference for k_{sl}/f_0 for ULES and LLES, respectively, normalized with the higher of k_{sl}/f_0 value between left and right. Thus, two relative features ($N_{FI_3} = 2$).
4) Left-right side ratio for $f_{0,L}/f_{0,R}$ for ULES and LLES, respectively. Same relative left-right ratio for $k_{sl,L}/k_{sl,R}$ was calculated. Thus, four relative features ($N_{FI_4} = 4$).
5) Up-down ipsilateral ratio for $f_{0,U}/f_{0,D}$ for left side and right side, respectively. Same relative up-down ratio for $k_{sl,U}/k_{sl,D}$ was calculated. Thus, four relative features ($N_{FI_5} = 4$).
6) Ratio k_{sl}/f_0 for all four channels separately. Thus, four relative features per subject ($N_{FI_6} = 4$).

Altogether, this feature group provided a total of twenty-six features, $N_{FI} = 26$.

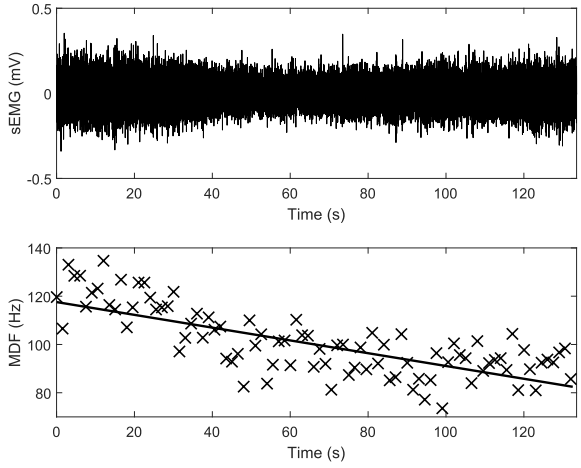

FIGURE 5. Raw sEMG signal (top) acquired with FREEEMG measurement system (section II-B). Linear regression (solid line) calculated LMSE-wise in respect to MDF values (crosses). Regression line slope corresponds to k_{sl} and intersection with y-axis corresponds to the initial frequency f_0.

F. OVERALL FEATURES SET

In this step, all calculated features are gathered to create a feature vector for each subject at hand. It is important to notice that each of fourteen primary features (Table 2)

are contributing to corresponding secondary features group (section IV-A), except for fatigue-related indices where only MDF feature is used. As a result, the number of variables per single observation is multifold increased. Accordingly, the number of features per secondary feature group is as follows:

1) Coordination measures: ten secondary features per single primary feature ($N_{Co} = 10$), altogether $N_{sg1} = N_{Co} \times 14 = 140$
2) Co-activation triggers: eight secondary feature per single primary feature ($N_{Ca} = 8$), altogether $N_{sg2} = N_{Ca} \times 14 = 112$
3) Trends: ten secondary features per single primary feature ($N_{Tr} = 10$), altogether $N_{sg3} = N_{Tr} \times 14 = 140$
4) Fatigue-related indices: twenty-six secondary features calculated directly from MDF, as presented in section IV-E, altogether $N_{sg4} = N_{FI} = 26$

On top of this, for coordination and co-activation features specifically, the autocorrelation-based features were also created. More concretely, before applying distance measures (coordination) or triggers detections (co-activation), (time)autocorrelation for each primary feature was calculated following the expression:

$$r_k = \frac{c_k}{c_0},\ c_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})\,(y_{t+k} - \bar{y}) \tag{16}$$

where k represents a lag (in our case a full shift across the whole sequence length was used), c_0 stands for the variance of the given time sequence y_t. This way, the number of secondary features for coordination and co-activation groups is doubled.

Additionally, for comparison and validation reasons, the overall secondary features set was expanded with single-valued representations of primary features, in simplistically. Namely, each primary feature time sequence (calculated directly from the raw sEMG signal as presented in section III) was contributing with only one value, and that is, the maximum value calculated on the whole sequence (mean, median, 95th, and 99th percentile values were also explored). Thus, a macroscopic representation of each primary feature was included, but at the cost of losing insight into time-related dynamics of myoelectric activity. This approach resulted with four additional features (one per channel) per each primary feature, altogether $N_P = 4 \times 14 = 56$ features.

As a last contribution to the overall feature set, the duration of the isometric static contraction (time-to-endurance, T_{EMG}) is added [6], [75], thus one feature per subject ($N_T = 1$).

Finally, the total number of features is the following sum $N_{sg,Tot} = 2 \times (N_{sg1} + N_{sg2}) + N_{sg3} + N_{sg4} + N_P + N_T = 727$.

V. FEATURE ANALYSIS AND SELECTION

A. PRIMARY FEATURES RELATIONSHIPS ANALYSIS

The initial set of fourteen primary features led to a large number of secondary features ($N_{Tot} = 727$). As the next step, the linear relationships were checked by calculating the correlation for each pair of primary features. For this correlation analysis, the intention was to preserve the aspect of time progression for the intrinsic characteristics of respective feature sequences. Therefore, for each primary feature sequence, a new vector was composed by stacking individual feature sequences for each subject, one on top of the other. Such vertically stacked vectors, for each feature, were then stacked column-wise into the matrix for all primary features at hand. The matrix composition is given by:

$$\begin{bmatrix} f_{1,1,1} & & f_{N_f,1,1} \\ f_{1,1,2} & \cdots & f_{N_f,1,2} \\ \vdots & & \vdots \\ f_{1,1,n_1} & & f_{N_f,1,n_1} \\ f_{1,2,1} & & \\ f_{1,2,2} & & f_{N_f,2,1} \\ \vdots & \cdots & f_{N_f,2,2} \\ f_{1,2,n_2} & & \vdots \\ \vdots & \vdots & \vdots \\ f_{1,N_s,1} & & f_{N_f,N_s,1} \\ f_{1,N_s,2} & & f_{N_f,N_s,2} \\ \vdots & \cdots & \vdots \\ f_{1,N_s,n_{N_s}} & & f_{N_f,N_s,n_{N_s}} \end{bmatrix} \tag{17}$$

where indices triplet (i, j, k) for each matrix element corresponds to i-th primary feature (N_f represents the total number of features), j-th subject s (N_s represents the total number of subjects) and k-th element in the time sequence of the respective primary feature with sequence length n_{N_s} for subject s (all feature sequences for same subject are of the same length). The feature vector for each subject is standardized ($\mu = 0$, $\sigma = 1$) before being added to the matrix to eliminate the effects of different absolute value ranges among the same features across different subjects. The correlation coefficient was calculated using Pearson correlation as given by (6). A correlation matrix (14×14) for each channel was obtained providing a valuable insight with Pearson correlation ρ and p-values using Student's t distribution for a transformation of the correlation [77]. Some level of correlation among all feature pairs was detected with p-values $p < 0.05$, i.e. not for a single pair of features it can be stated with confidence >95% that $\rho = 0$, although, for most of the pairs, it was relating to low-level correlation. However, a very strong correlation, $\rho^2 > 0.8$, has been observed for the following feature subsets across all channels (Fig. 6), and irrespective of the subjects' health condition state (HS vs. CLBP vs. RLBP).

1) MAV, IEMG, LD, RMS, VAR
2) SSC, PE

B. FEATURE SELECTION FOR SECONDARY GROUPS

Neighborhood component analysis (NCA) was selected as the main technique for feature selection in our study [78]. NCA is a supervised non-parametric feature selection approach capable of removing irrelevant features, thus effectively

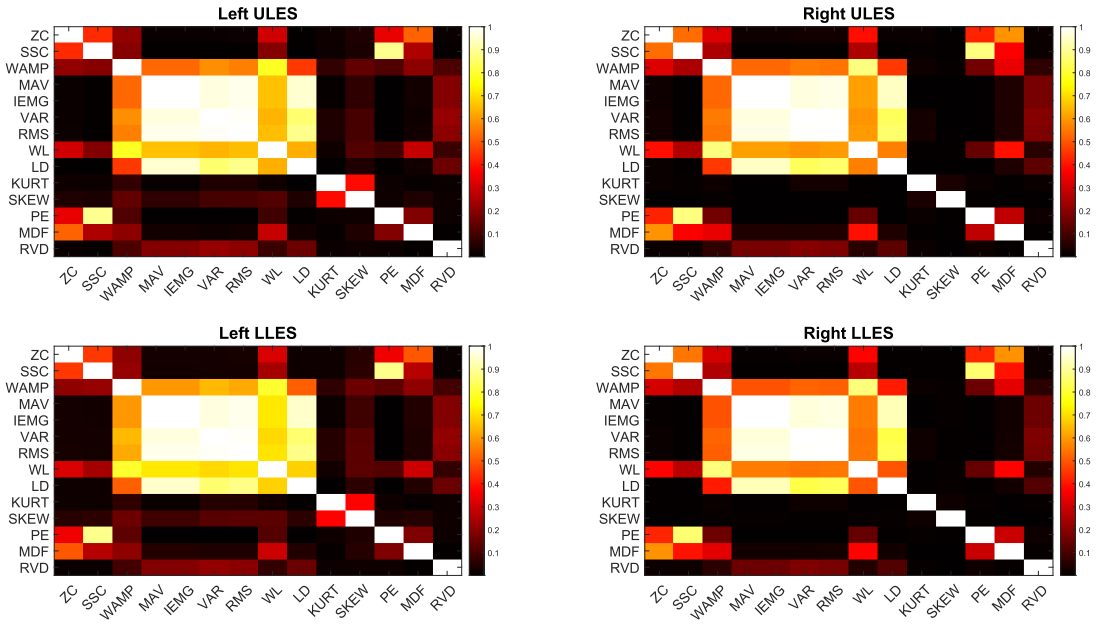

FIGURE 6. Correlation matrix for all fourteen primary features (ZC, SSC, WAMP, MAV, IEMG, VAR, RMS, WL, LD, KURT, SKEW, PE, MDF, RVD) for a set of LBP subjects (CLBP and RLBP) only, and for each sEMG channel, left and right ULES (top), and left and right LLES (bottom). Correlation values are shown as squared Pearson correlation coefficient, ρ^2.

reducing the high-dimensional data while keeping the selected feature vectors in the original vector space. This is unlike some other commonly used feature reduction or feature projection techniques, most notably principal component analysis (PCA). Such an approach opened up the possibility of getting a better insight into understanding which features and properties have the key impact on LBP detection and differentiation among subjects.

NCA method, implicitly, throughout the procedure of maximizing the objective function, measures the average leave-one-out classification accuracy and sets respective weights $\boldsymbol{w}$) for individual features [79]. The classification method used behind is a variant of the 1-NN classifier. The aim is to find such weights vector $\boldsymbol{w}$ that will optimize the nearest neighbor classification accuracy. In the context of weights vector $\boldsymbol{w}$, a weighted distance between two samples is defined with

$$D_{\boldsymbol{w}}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = \sum_{l=1}^{d} w_l^2 \left|x_{il} - x_{jl}\right|. \tag{18}$$

To maximize the leave-one-out classification accuracy by employing the nearest neighbor, a classification accuracy function in a differential form is required. Knowing that the leave-one-out form is not a differential function *per se*, we are looking into the approximation of the probability distribution function with the following expression

$$p_{ij} = \begin{cases} \dfrac{\kappa\left(D_{\boldsymbol{w}}(\boldsymbol{x_i}, \boldsymbol{x_j})\right)}{\sum_{k \neq i} \kappa\left(D_{\boldsymbol{w}}(\boldsymbol{x_i}, \boldsymbol{x_k})\right)}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \tag{19}$$

where $\kappa(z) = exp(-z/\sigma)$ is a kernel function of kernel width σ. Kernel function was chosen in such a way to secure that for small distances $D_{\boldsymbol{w}}(\boldsymbol{x_i}, \boldsymbol{x_j})$, the resulting probability p_{ij} (to have sample $\boldsymbol{x_i}$ choose $\boldsymbol{x_j}$ as a reference point) is set high. Two characteristic edge cases for kernel width σ can be observed:

- $\sigma \to 0$ where only the closest neighbor can be selected as a reference point
- $\sigma \to +\infty$ where all points have equal chance to be selected (except $\boldsymbol{x_i}$ point)

Following (18) and (19), the probability to have point $\boldsymbol{x_i}$ correctly classified is

$$p_i = \sum_j y_{ij} p_{ij} \tag{20}$$

where $y_{ij} = 1$ if and only if $y_i = y_j$, otherwise $y_{ij} = 0$. Therefore, the approximate classification accuracy can be written as

$$\xi\left(\boldsymbol{w}\right) = \frac{1}{N} \sum_i p_i = \frac{1}{N} \sum_i \sum_j y_{ij} p_{ij} \tag{21}$$

where for $\sigma \to 0$, $\xi\left(\boldsymbol{w}\right)$ becomes exact leave-one-out classification accuracy. Furthermore, to enable the feature selection (and prevent classification model overfitting), a new regularization factor is introduced resulting in the subsequent objective function:

$$\xi\left(\boldsymbol{w}\right) = \sum_i \sum_j y_{ij} p_{ij} - \lambda \sum_{l=1} w_l^2 \tag{22}$$

where λ stands for the regularization parameter which can be further fine-tuned throughout an iterative process with cross-validation. Furthermore, it is also good to note that only one

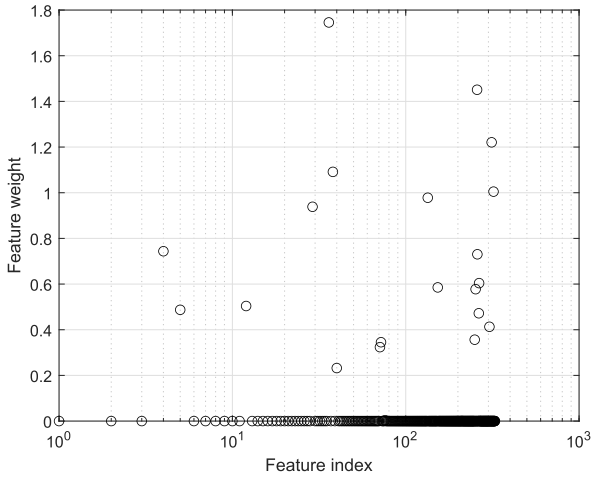

FIGURE 7. Example of feature weights for each secondary feature as part of the NCA feature selection procedure for the submitted secondary features set based on SSC, WAMP, VAR, ZC, RVD, and PE primary features. The secondary feature set consisted of 327 features. Only features with feature weights >0.02 were included in the resulting subset of features. Twenty features were selected this way.

regularization parameter λ is used for all the weighted factors w_n, thus bringing less complexity in finding an adequate value.

Now, when having a differentiable (approximate) function $\xi(\boldsymbol{w})$, we are searching for such $\boldsymbol{w}$ vector for which $\frac{\delta\xi(\boldsymbol{w})}{\delta w_l}$ will provide maximum value. This search is done by employing some of the iterative numeric gradient optimization methods, like the proposed gradient ascent technique [79], [80]. It is shown that the given optimization procedure, by including the regularization parameter, results in many weighted factors $w_n \rightarrow 0$. Features with such weights are excluded, thus effectively reducing the number of features significantly. In this study, a threshold of 0.02 was applied as an exclusion criterion. The remaining features are forming the selected features subset for subsequent classification steps.

An example of NCA feature selection based on resulting feature weights, is given with Fig. 7, where resulting 20 features were selected out of the initial 327.

VI. CLASSIFICATION MODELS AND VALIDATION PROCEDURE

In this section, we are employing different classification models in order to verify our proposed method based on sEMG with LBP-specific features that were previously constructed. The intention was to establish a meaningful set of features that can be correlated with actual subjects' health conditions expressed via myoelectric activity, triggered by endurance exercise and related neuromuscular underlying processes, with or without the presence of LBP.

As presented in Table 1, our data set consisted of ninety-one subjects, divided into three groups (HS, CLBP, RLBP), with a share of 40.7%, 31.9%, and 27.4% among groups, respectively. Demographic and biometric statistics do not point to any significant deviation or differences among groups when considering age, height, weight, and BMI.

A. CLASSIFICATION MODELS

Several different groups of classifier types and classification models were examined in combination with our constructed feature sets. We used *Classification Learner* application as part of Matlab R2020b (The Mathworks, Inc., Natick, Massachusetts) software package [81] to prepare data, select the classification models, conduct classification procedures and validate respective models. We have chosen predefined classification models *as-is* without any additional parameters optimization or fine-tuning steps. Thus, the goal was to primarily validate the potential of our proposed approach with LBP domain-related features constructed as previously discussed. Matlab classification models in this study, used *out-of-the-box*, can be grouped into the following:

- Decision trees (DT) with fine, medium and coarse splits. Strong foundations for this approach and its popularity were established by L. **B**reiman, J. **F**riedman, R. **O**lshen i C. **S**tone (BFOS group) [82].
- Discriminant analysis: linear (LDA) and quadratic (QDA). Based on Fisher's discriminant analysis [83], [84].
- Logistic regression (LR) [85]
- SVM with different kernel types (linear, quadratic, cubic, Gaussian) and with different scales for Gaussian kernels (fine, medium and coarse). [86], [87].
- *k*-nearest neighbor (*k*NN) with fine, medium and coarse resolution depending on parameter (k) and different distance metrics (Euclidean, cosine or cubic), with or without a weighted distance [88], [89].
- Ensemble classifiers based on *bagging* [90], *boosting* [91]–[93] and *random subspace* techniques [94], [95].

A given list resulted with twenty-three different classification model variants used to validate selected feature sets and proposed method overall as listed in Tables 3, 5, 6. Neural networks-based classifiers were not in focus in this study. It was primarily due to such classifiers being characterized by their black-box nature and challenges in training models when sample sizes are somewhat limited (as it is more pertained to this specific LBP domain).

1) DECISION TREES

DT algorithms are based on the recursively repeated splits of data set by applying the selected criterion that maximizes the separation of the data, which in turn creates a tree-like structure [82].

A classic CART algorithm [82] was employed with the predefined splitting criterion (Gini's diversity index) and splitting resolutions (namely, fine, medium, and coarse with 100, 20, 1, as a maximum number of splits, respectively).

DTs were of interest for our study as enabling insight into the models that can easily be expressed as a set of rules (resulting models are not black-boxes).

2) DISCRIMINANT ANALYSIS

LDA and QDA are classification and dimensionality reduction techniques that are developed based on Fisher's linear discriminant. These methods tend to express a dependent variable (sample class) as a linear combination of other independent features. The classification objective is to find such a linear combination that best separates two or more classes [84].

Classification models employed in this study were utilizing a full covariance matrix structure as a predefined value for both LDA and QDA classifiers (as compared to the diagonal-only option).

3) LOGISTIC REGRESSION

Logistic regression is a type of statistical model that employs a logistic function (sigmoid function) to construct probability models for binary dependent variables. Such probability models can be turned into classifiers by introducing a threshold value where input values (e.g. feature vectors) with a probability above the threshold are classified as one class, and if below, as the other class.

Model parameter values [85] were implicitly estimated by *Classification Learner* application [81].

4) SUPPORT VECTOR MACHINES

SVM is one of the most popular and most used machine learning and classification techniques. It was shown to be very robust and among the most accurate methods based on good mathematical and theoretical foundations. Some of the key strengths are the ability to learn on a limited number of training samples and to successfully deal with high feature attributes dimensionality (thus, performance independent on a number of dimensions) [73].

SVM models are generally characterized by the following parameters: kernel type, kernel size, and separability margin type (hard- or soft-margin). Some of the kernel functions K types can be defined with:

- $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + 1)^p$, where by selecting value for parameter p ($p = 1$, 2 or 3), a linear, quadratic or cubic kernel is defined, respectively.
- $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\|x - y\|^2 \big/ 2\sigma^2\right\}$, where the exponential function represents the radial basis, i.e. Gaussian kernel function

In this setup, auto-selected values, as per [81], for kernel widths (Gaussian kernel function) were applied for three different resolution variants: fine, medium, and coarse resolution. Soft-margin, as a predefined separation criterion, was applied to all SVM classifiers variants employed [86], [87].

5) *k*-NEAREST NEIGHBOR

*k*NN is a simple classification algorithm that can perform well in many scenarios, including multiclass classification cases. Key elements that determine the outcome of such a nearest neighbor approach are (A) selection of value k, (B) distance or similarity measure, and (C) the structure of the labeled set. The choice of k, if set too small, can result in models sensitive to noise points. On the other hand, for large k, a risk of including too many points from other classes in the neighborhood is imminent [73]. As for distance measures (B), a key principle is to engage such a metric for which a smaller distance between two objects implies a greater likelihood of having the same class.

In this study, predefined k values 1, 10, and 100 were tested, resulting in fine, medium, and coarse resolution, respectively (with predefined Euclidean distance used). Also, two other distance metrics were exploited, cosine and cubic, with medium resolution ($k = 10$). Finally, a variant with square-inverse weighted distance measure was employed as well (on Euclidean distance), instead of simple majority voting (counting) approach.

6) ENSEMBLE LEARNING

In this study boosted trees classifier, based on AdaBoost algorithm was utilized [92]. Additionally, another variant of boosting ensemble classifier was also employed, namely, **R**andom **U**nder**s**ampling (RUS) boosting trees [95]. A key advantage is an effectiveness at classifying imbalanced data, in cases when some class in the training data set has many fewer members than another.

Also, the bagged trees classifier was used where, in a nutshell, it is a method where multiple versions of a classifier are generated by bootstrap replicates (i.e. sampling with replacement) of the learning set and using these as new learning sets [90]. The overall classifier is constructed by combining classification results in a majority voting manner.

Further, random subspace methods were employed where an ensemble classifier was constructed by a series of (weak) learners that operate on a subset (thus *subspace*) of randomly selected features [94]. Linear discriminant classifier and *k*-nearest neighbor were employed as (weak) learners. Classification prediction outcome was constructed by taking an average of the score prediction of the weak learners, and classifying the category with the highest average score [81].

As in the previous cases, predefined out-of-the-box parameter values were applied, where for all ensemble classifiers the number of learners was set to 30, with a learning rate of 0.1. Additionally, for boosted algorithms (boosted trees and RUSBoosted trees), the maximum number of splits was set to 20, whereas for bagged trees it was set to 90. Gini's diversity index, as a default split criterion, was used in both cases.

B. CROSS-VALIDATION

Feature selection (based on NCA) and models verification (twenty-three classifiers) procedures were performed on a full data set in an M-fold cross-validation manner. Meaning, no separate untouched test set was used, but the data set at

hand was divided into M nonoverlapping subsets of equal size, where the prediction model was trained on $M-1$ folds combined together, whereas the prediction accuracy (i.e. error) was estimated on the remaining samples (single fold). Cross-validation was opted to make full use of all LBP patient samples having in mind the specific LBP domain challenges and the expected heterogeneity among the patients. This approach is relatively common in cases when the number of samples in the data set is relatively small. However, this approach introduces certain selection bias that generally can result in a too-optimistic estimate of the prediction error rate [96], [97].

For the NCA feature selection procedure, the number of folds was equal to the number of samples in the data set, thus leave-one-out cross-validation was performed. Leave-one-out cross-validation is considered to be nearly unbiased, but with potentially high variance. For the verification of models utilizing twenty-three classifiers, a cross-validation procedure with ten-folds ($k_{fold} = 10$) was performed for every single classification. Further, each classification with cross-validation was repeated ten times ($N_{iter} = 10$). Finally, all results were averaged. In this way, more statistical confidence by reducing the effects of results' variations was introduced.

Alongside classification accuracy (ACC), the values for sensitivity (recall, RC) and precision (positive predictive value, PPV) were also calculated in the same manner as a result of multiple iterations and averaging.

Additionally, each classification experiment was conducted fully independent, thus different classification comparisons (Table 5 and Table 6) that were employing the same feature sets and models (e.g. feature sets relating to Top 5+PE), could have resulted with slightly different results in different experiments, due to statistical variation, but in the same time giving the notion of results consistency.

VII. RESULTS

In previous sections II−VI, the proposed method for raw feature extraction, novel feature construction and feature selection were presented. The following section is presenting results with the aim to: (A) verify the choice of features selected and (B) apply the feature construction outcomes to the classification models to detect LBP patients and consequently make differentiation among subject groups at hand (HS, CLBP, RLBP).

Features verification process (A) was established by performing a series of classifications and cross-validations with comparisons, through two-class classifiers. Therefore, our data set (ninety-one subjects) was initially divided into two groups, namely, healthy subjects (HS) and LBP patients (CLBP and RLBP combined), with thirty-seven and fifty-four subjects, respectively (Table 1). Following feature validation outcomes in (A), we performed a series of two-class classifications (B) among the three groups at hand, separately. More concretely, two-class classifications were performed for the following differentiation among groups: HS vs. CLBP, HS vs. RLBP, and CLBP vs. RLBP. As a basic step, the overall HS vs. LBP differentiation was performed, where general separability between LBP patients and non-LBP subjects was tested.

A. INDIVIDUAL FEATURES VERIFICATION

In this procedure step, we wanted to verify the contribution of each individual primary feature to the overall feature set, as defined in section IV-F, by relating these findings with the earlier performed correlation analysis pointing to certain collinearity within the subset of features (Fig. 6).

The verification across all primary features and all classifiers was conducted for HS vs. LBP two-class classification, as follows:

1) For each primary feature, the corresponding secondary features set was calculated (based only on that single primary feature)
2) NCA procedure for features selection was performed (thus, reducing the number of features we are working with)
3) Each of twenty-three classification models was employed by taking previously NCA selected features set as an input
4) Cross-validation procedure with ten folds ($k_{fold} = 10$) was performed, as presented in section VI-B

Results for this intermediate verification step are presented in Table 3.

For each primary feature, an averaged classification accuracy rank was calculated across all classifiers employed to provide an insight into how well the respective feature resonated with the given secondary feature groups modeling. These ranks are given in Table 4.

SSC performed best for nine classifiers (especially for decision trees), whereas PE performed best for eleven classifiers (especially for SVM variants), out of twenty-three classifiers (and variants) employed. Among all classifiers and all features, SSC provided the single best classification accuracy of 0.88 for kNN classifier with fine resolution ($k = 1$). These results were reflected in the calculated average ranks where SSC and PE were labeled as the best and second-best, respectively.

Taking into account the classification results, the respective assigned ranks, as well as the correlation insights obtained in section V-A, an initial subset of five primary features was selected: SSC, WAMP, VAR, ZC, RVD. Features that were demonstrating high correlation (MAV, IEMG, LD, RMS), or limited classification contribution (MDF, WL, KURT, SKEW), were removed from the set. The contribution of PE was analyzed and consequently was added to the selected set, despite exhibiting a high correlation with SSC. PE demonstrated the highest median values (0.81) across all classifiers, compared to 0.79 for SSC, and the 2nd best average rank result (Table 4). Therefore, PE could not have

TABLE 3. Two-class classification (HS vs. LBP) accuracy results (with ten-fold cross-validation) with secondary features sets created based on each single primary feature (Table 2) only, from left to right, across all classifiers. Classifiers employed are decision trees (DT), discriminant analysis (LDA and QDA), logistic regression (LR), support vector machines (SVM), nearest neighbor (*k*NN), and ensemble methods (bagged, boosted, subspace, random undersampling - RUS), from top to bottom. For each classifier, the primary feature providing the highest accuracy is shown in boldface. Median (Median) and maximum (Max) values for each individual primary feature across all classifiers are shown as well.

Classification model	ZC	SSC	WAMP	MAV	IEMG	VAR	RMS	WL	LD	KURT	SKEW	PE	MDF	RVD
Fine DT	0.69	**0.78**	0.73	0.72	0.69	0.75	0.73	0.63	0.69	0.67	0.64	0.71	0.70	0.67
Medium DT	0.72	**0.79**	0.73	0.72	0.68	0.76	0.75	0.61	0.70	0.68	0.66	0.72	0.70	0.69
Coarse DT	0.67	**0.78**	0.72	0.70	0.70	0.77	0.76	0.70	0.69	0.63	0.66	0.70	0.70	0.67
LDA	0.76	0.79	0.77	0.76	0.76	0.78	0.75	0.75	0.76	0.71	0.75	**0.81**	0.71	0.78
QDA	0.78	**0.79**	0.74	0.69	0.69	0.79	0.68	0.75	0.78	0.72	0.68	**0.79**	0.68	0.73
LR	0.75	**0.79**	0.76	0.76	0.76	0.77	0.75	0.75	0.77	0.70	0.74	0.78	0.71	**0.79**
Linear SVM	0.75	0.78	0.77	0.74	0.73	0.77	0.77	0.75	0.75	0.73	0.75	**0.83**	0.71	0.78
Quadratic SVM	0.78	0.84	0.85	0.71	0.72	0.76	0.76	0.70	0.78	0.73	0.72	**0.86**	0.70	0.79
Cubic SVM	0.82	0.77	0.76	0.67	0.69	0.79	0.74	0.60	0.78	0.68	0.63	**0.83**	0.77	0.74
Fine Gaussian SVM	**0.74**	0.66	0.66	0.59	0.59	0.71	0.58	0.65	0.59	0.59	0.71	0.63	0.62	0.66
Medium Gaussian SVM	0.81	0.84	0.84	0.71	0.72	0.75	0.80	0.73	0.79	0.78	0.72	**0.86**	0.72	0.79
Coarse Gaussian SVM	0.79	0.80	0.77	0.67	0.68	0.76	0.77	0.76	0.71	0.70	0.75	**0.82**	0.73	0.78
Fine kNN	0.74	**0.88**	0.84	0.72	0.73	0.75	0.74	0.67	0.66	0.74	0.62	0.79	0.78	0.80
Medium kNN	0.77	0.81	0.81	0.72	0.72	0.78	0.78	0.74	0.74	0.75	0.73	**0.85**	0.73	0.74
Coarse kNN	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59
Cosine kNN	0.77	0.80	0.82	0.73	0.73	0.75	0.73	0.77	0.75	0.73	0.72	**0.83**	0.74	0.80
Cubic kNN	0.77	0.81	**0.82**	0.72	0.71	0.77	0.76	0.74	0.75	0.75	0.74	0.81	0.72	0.73
Weighted kNN	0.76	0.84	**0.86**	0.74	0.74	0.78	0.78	0.71	0.73	0.75	0.69	**0.86**	0.78	0.82
Boosted Trees[a]	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.43	0.41	0.41	0.52	0.41	0.41	0.41
Bagged Trees	0.75	**0.81**	0.79	0.72	0.71	0.76	0.76	0.70	0.77	0.72	0.69	0.79	0.77	0.76
Subspace Discriminant	0.77	0.81	0.75	0.73	0.74	0.76	0.77	0.78	0.73	0.73	0.76	**0.82**	0.73	0.77
Subspace kNN	0.75	**0.83**	0.82	0.72	0.72	0.73	0.75	0.71	0.70	0.77	0.67	0.82	0.78	0.78
RUS Boosted	0.71	**0.79**	0.74	0.70	0.71	0.73	0.72	0.65	0.76	0.68	0.66	0.75	0.74	0.69
Median	0.75	0.79	0.77	0.72	0.71	0.76	0.75	0.71	0.74	0.72	0.69	**0.81**	0.72	0.76
Max	0.82	**0.88**	0.86	0.76	0.76	0.79	0.8	0.78	0.79	0.78	0.76	0.86	0.78	0.82

[a] Underperforming classifier(s) with classification accuracy <50%, for most of the feature sets.

TABLE 4. Sorted average rank list where features with best average classification rank across all classifiers are set on top.

Primary feature	Average rank
SSC	2.38
PE	2.62
WAMP	4.10
VAR	5.05
ZC	6.17
RVD	6.29
RMS	7.40
LD	8.19
MDF	9.45
WL	10.07
MAV	10.45
IEMG	10.57
KURT	10.71
SKEW	11.55

been ignored despite its correlation with the best ranked SSC feature.

B. FEATURE SETS CONFIRMATION

Altogether insights from the previous steps were exploited to define the primary features subset where the top five primary features (SSC, WAMP, VAR, ZC, RVD) as per section VII-A, with additionally including PE, were selected. This feature set was labeled as Top 5+PE.

Further, it was also of interest to examine the contribution of each secondary features group and enable inference of such contribution to the classification results. Consequently, such insights were meant to foster discussion about the potential of the proposed approach and the justification behind it. Therefore, a separate classification procedure was performed where each individual secondary feature group (Coordination, Co-activation, Trends, Fatigue indices as per section IV-F) was chosen as an input into the classifier models (again, across the same twenty-three classifier variants). The selected Top 5+PE set was employed (SSC, WAMP, VAR, ZC, RVD + PE) for constructing each secondary group.

Results are presented in Table 5. It can be observed that the Coordination features group exhibited overall best accuracy, among secondary feature groups, with a median value of 0.81 across all classifiers, closely followed by the Co-activation features group with 0.80 median value. Trends, Fatigue indices, and Primary groups exhibited notably less successfully, with the Primary group exhibiting even somewhat better than the remaining two groups. Overall best accuracy result was achieved for the Complete set (all secondary groups combined), with 0.85 median value. Additionally, it can be noticed that Coordination groups performed significantly better for decision trees type of classifiers (accuracy ranging from 0.82 to 0.85) compared to other feature sets, whereas the Complete set significantly outperformed the SVM classifiers with accuracy results >0.90.

C. COMPARISON AMONG SUBJECTS GROUPS

One of the main tasks was to provide insights into the proposed method's capability of dealing with differentiation

TABLE 5. Two-class classification (HS vs. LBP) accuracy results (with ten-fold cross-validation) for different secondary feature groups (section IV-A). The Top 5+PE subset of primary features is employed. Primary features (Primary max), leftmost group, together with a complete feature set consisting of all secondary groups (Complete set), rightmost, are validated and shown as well. NCA feature selection procedure was performed for all feature groups with the number of components given by N_{NCA}. The best performing feature group for the given classifier is shown in boldface. Classifiers with any of the accuracy results ≥90% are shown in boldface. Median (Median) and maximum (Max) values for each secondary feature group across all classifiers are shown as well.

Classification model	Primary max $N_{NCA} = 6$	Fatigue indices $N_{NCA} = 7$	Trends $N_{NCA} = 7$	Coordination $N_{NCA} = 11$	Co-activation $N_{NCA} = 12$	Complete set $N_{NCA} = 20$
Fine DT	0.66	0.63	0.72	**0.84**	0.68	0.80
Medium DT	0.69	0.65	0.71	**0.85**	0.68	0.78
Coarse DT	0.64	0.56	0.72	**0.82**	0.68	0.76
LDA	0.77	0.58	0.76	0.78	0.82	**0.84**
QDA[a]	N/A	**0.80**	0.76	**0.80**	0.75	0.75
LR	0.76	0.60	0.77	0.80	0.80	**0.84**
Linear SVM	0.76	0.62	0.77	0.80	0.84	**0.86**
Quadratic SVM	0.79	0.65	0.81	0.82	0.80	**0.94***
Cubic SVM	0.78	0.61	0.79	0.82	0.88	**0.94***
Fine Gaussian SVM	0.61	**0.75**	0.70	0.67	0.59	0.59
Medium Gaussian SVM	0.78	0.63	0.78	0.82	0.84	**0.93***
Coarse Gaussian SVM	0.74	0.59	0.75	0.77	0.77	**0.83**
Fine kNN	0.79	0.66	0.73	**0.91***	0.85	0.90*
Medium kNN	0.80	0.58	0.79	0.81	0.81	**0.84**
Coarse kNN	0.59	0.59	0.59	0.59	0.59	0.59
Cosine kNN	0.79	0.61	0.79	0.81	0.77	**0.82**
Cubic kNN	0.81	0.65	0.80	0.81	0.80	**0.86**
Weighted kNN	0.79	0.70	0.79	0.84	0.82	**0.85**
Boosted Trees[b]	0.41	0.43	0.41	0.41	0.41	0.41
Bagged Trees	0.79	0.66	0.79	0.80	0.80	**0.84**
Subspace Discriminant	0.77	0.55	0.76	0.79	0.81	**0.86**
Subspace kNN	0.71	0.63	0.78	**0.90***	0.84	**0.90***
RUS Boosted	0.75	0.63	0.74	**0.80**	0.72	0.78
Median	0.77	0.63	0.76	0.81	0.80	**0.84**
Max	0.81	0.80	0.81	0.91	0.88	**0.94**

*Classification results with accuracy ≥90%.
[a]Some of QDA classification models failed due to singular covariance matrix for one of the classes.
[b]Underperforming classifier(s) with classification accuracy <50%.

among subjects groups, thus going a step beyond only separating the healthy subjects (HS) from patients (LBP). Therefore, another classification iteration was conducted with an aim to determine the success of group differentiation for (I) HS vs. LBP, (II) HS vs. RLBP, (III) HS vs. CLBP, and (IV) CLBP vs. RLBP pairs. Patients groups, namely LBP, RLBP, and CLBP, in comparison with HS, were labeled as a positive class (1), whereas HS was labeled as a negative class (0). In comparison between CLBP and RLBP, RLBP was labeled a positive class (1). ACC, PPV and RC as measures of classification success were also calculated and given in Table 6. For each feature group, detailed insight into the structure of NCA components was also given, thus providing a tool for a deeper inference into the underlying muscle activity processes by means of sEMG (Table 7).

Overall, noticeably, the best classification results were achieved for HS vs. RLBP differentiation with classification accuracy ranging from 0.93 to 0.98 for all classifiers except coarse *k*NN, boosted trees and RUS boosted with the accuracy of 0.60, 0.60, and 0.78, respectively. Moreover, for HS vs. RLBP differentiation, sensitivity across all classifiers was in the range 0.95 to 1.00, where precision for all classifiers, except coarse *k*NN, boosted trees and RUS boosted, was in the range 0.90 to 0.99.

The second-best classification results were obtained for HS vs. LBP classification with the best accuracy achieved for SVM variants, namely, for quadratic, cubic and medium Gaussian kernel classifiers, with 0.93, 0.94, 0.94, respectively. Sensitivity results >0.95 were achieved for quadratic and cubic SVM, 0.96 and 0.97, respectively. The highest result for sensitivity was obtained with boosted trees (100%), but, similarly as for fine Gaussian SVM and coarse *k*NN, resulted in 0% sensitivity and undefined precision. This was due to having all samples labeled with either positive or negative class, thus these results were considered fallacious.

For HS vs. CLBP differentiation, the best accuracy results of 0.90 were obtained by fine resolution *k*NN and subspace *k*NN. Fine *k*NN was also the only classifier exhibiting precision >0.90, whereas, for sensitivity, multiple classifiers exhibited values >0.90, with fine Gaussian SVM, coarse *k*NN and boosted trees moreover exhibiting sensitivity of 100%, but at the expense of low accuracy and precision. Cubic SVM, fine *k*NN, and subspace *k*NN provided a good balance of classification success for HS vs. CLBP across all measured values (ACC, PPV, RC).

The least performing classification differentiation pair was for CLBP vs. RLBP. None of the classifiers achieved accuracy >0.90, where the best classification accuracy of 0.89 was achieved for fine *k*NN. The same classifier was the only one to exhibit a precision >0.90, namely 0.92. Apart from the low-performing classifiers (fine Gaussian SVM, coarse *k*NN, and boosted trees), the best-performing

TABLE 6. Two-class classifier (with ten-fold cross-validation) results for accuracy (ACC), precision (PPV), sensitivity (RC) are shown. A series of classification differentiation pairs are (I) HS vs. LBP, (II) HS vs. RLBP, (III) HS vs. CLBP, (IV) CLBP vs. RLBP for healthy (HS) subjects, subject with radiculopathy (RLBP), chronic low back pain patients (CLBP), and LBP as a joint group for RLBP and CLBP patients. ACC, PPC, and RC results ≥90% are shown in boldface. Highlighted rows show best performing classifiers. Median (Median) and maximum (Max) values for each classification pair for ACC, PPV and RC are calculated.

Classification model	ACC				PPV				RC			
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
Fine DT	0.77	**0.94**	0.74	0.68	0.72	**0.94**	0.76	0.71	0.74	**0.95**	0.79	0.68
Medium DT	0.78	**0.94**	0.73	0.69	0.73	**0.94**	0.75	0.72	0.74	**0.96**	0.79	0.70
Coarse DT	0.73	**0.93**	0.69	0.69	0.69	**0.94**	0.72	0.71	0.63	**0.95**	0.75	0.70
LDA	0.85	**0.98**	0.75	0.74	0.77	**0.97**	0.77	0.78	0.88	**1.00**	0.78	0.73
QDA[a]	0.75	**0.96**	0.78	N/A[a]	0.78	**0.97**	0.80	N/A[a]	0.52	**0.96**	0.80	N/A[a]
LR	0.83	**0.95**	0.75	0.75	0.76	**0.95**	0.77	0.77	0.86	**0.97**	0.78	0.76
Linear SVM	0.87	**0.96**	0.76	0.67	0.81	**0.97**	0.77	0.71	0.89	**0.97**	0.81	0.67
Quadratic SVM	**0.93**	**0.97**	0.84	0.83	0.89	**0.97**	0.86	0.83	**0.96**	**0.98**	0.86	0.85
Cubic SVM	**0.94**	**0.97**	0.88	0.82	0.89	**0.97**	0.89	0.84	**0.97**	**0.98**	**0.90**	0.82
Fine Gaussian SVM[b]	0.59	**0.93**	0.56	0.54	N/A[b]	**0.90**	0.56	0.54	0.00[b]	**1.00**	**1.00**	**1.00**
Medium Gaussian SVM	**0.94**	**0.98**	0.85	0.77	**0.93**	**0.97**	0.85	0.78	**0.91**	**0.99**	0.89	0.81
Coarse Gaussian SVM	0.82	**0.98**	0.74	0.64	0.82	**0.97**	0.69	0.60	0.72	**1.00**	**0.99**	**0.96**
Fine kNN	**0.90**	**0.98**	**0.90**	0.89	0.85	**0.97**	**0.91**	**0.92**	**0.92**	**1.00**	**0.92**	0.87
Medium kNN	0.84	**0.98**	0.82	0.74	0.75	**0.97**	0.78	0.73	**0.90**	**1.00**	**0.94**	0.81
Coarse kNN[b]	0.59	0.60	0.56	0.54	N/A[b]	0.60	0.56	0.54	0.00[b]	**1.00**	**1.00**	**1.00**
Cosine kNN	0.83	**0.98**	0.82	0.74	0.74	**0.97**	0.79	0.75	0.88	**1.00**	**0.92**	0.79
Cubic kNN	0.85	**0.98**	0.78	0.74	0.78	**0.97**	0.76	0.73	**0.90**	**1.00**	**0.91**	0.80
Weighted kNN	0.85	**0.98**	0.80	0.80	0.77	**0.97**	0.77	0.79	**0.90**	**1.00**	**0.92**	0.84
Boosted Trees	0.41	0.60	0.56	0.54	0.41	0.60	0.56	0.54	**1.00**	**1.00**	**1.00**	**1.00**
Bagged Trees	0.86	**0.95**	0.79	0.74	0.82	**0.95**	0.77	0.74	0.85	**0.98**	0.88	0.79
Subspace Discriminant	0.85	**0.97**	0.77	0.74	0.80	**0.95**	0.76	0.76	0.84	**1.00**	0.85	0.77
Subspace kNN	**0.90**	**0.96**	**0.90**	0.86	0.84	**0.99**	0.87	0.85	**0.94**	**0.95**	**0.97**	**0.90**
RUS Boosted	0.79	0.78	0.74	0.69	0.71	0.74	0.75	0.71	0.80	**0.98**	0.81	0.73
Median	0.84	0.96	0.77	0.74	0.78	0.97	0.77	0.74	0.88	0.99	0.89	0.81
Max	0.94	0.98	0.90	0.89	0.93	0.99	0.91	0.92	1.00	1.00	1.00	1.00

[a] Some of QDA classification models failed due to singular covariance matrix for one of the classes.
[b] Classification failed due to all subjects being classified to only one (negative) class label.

sensitivity was demonstrated by fine *k*NN and subspace *k*NN, with 0.87 and 0.90, respectively.

VIII. DISCUSSION

A. ABOUT FEATURE MODELING

The main focus of this study was to model such feature sets that enable satisfactory detection of LBP patients, alongside a possibility to enable subgrouping within heterogenous LBP groups and with the ability to interpret results from a medical point of interest. Features were constructed with a motivation to reflect certain neuromuscular and motor control changes pertained to LBP, as discussed and elaborated in a number of studies [4], [28], [37], [75], [98]–[101]. Measures of LBP coordination, co-activation, muscle fatigue, and trends tracked throughout the isometric exercise duration were selected as the key indicators for LBP classification modeling in this study.

The coordination aspect was taken into account as different studies reported asymmetric bilateral (left-right) muscle activation patterns to be observed for LBP patients in the lumbar region during functional tasks [67], [102], compared to healthy subjects where such lumbar muscle coordination exhibited more symmetric patterns. Additionally, Liu *et al.* proposed modeling the activity coordination network between lumbar muscles based on flexion-extension tasks performed in sEMG electrode matrix setup. Results demonstrated that healthy subjects clearly exhibited globally symmetric patterns between the left and right side of sEMG channels, whereas the LBP patients group showed the loss of global symmetric patterns. However, Revees *et al.* [103] compared varsity athletes with and without low-back injury history and reported no group differences in the imbalance between sides, during an isometric trunk extension exercise. Some earlier studies by Roy *et al.* [75] also reported that left-right differences were present in both LBP and control groups, recorded during isometric contraction exercise, thus not being specific to one group particularly. The presented results demonstrated certain ambiguities in reported findings, likely to come due to different setups as well (e.g. isometric exercise or functional tasks, athletes or the general population, few sEMG channels or electrode arrays, etc.), with no univocal conclusion about how the coordination patterns contribute to the groups' differentiation. Hence, the idea to include coordination aspects into consideration, and to further investigate such aspects of muscle activation imbalances between sides, seemed plausible.

In this study, primarily the bilateral relations for ULES and LLES were analyzed (section IV-B), thus not considering contralateral imbalances as in some other studies [99]. Moreover, our proposed coordination measure was based on the similarity measures calculated intra-subject between left and right muscle sites, by pairing time-aligned time epochs ($L = 1000$ ms, $s = 50$ ms) between muscle sites throughout the whole exercise duration. This approach is

different from studies where measures (e.g. MDF, RMS) were calculated for each muscle site separately, and afterward, such single-valued measures, were compared to establish a macroscopic significance of existing imbalance between sides [75], [99], [103].

In order to capture more subtle co-activation patterns across all four muscle sites that were examined, an additional measure of co-activation triggers score was introduced (section IV-C). When considering movement-related LBP differentiation based on the functional type of tasks (e.g. flexion-extension) it is shown that there is a significant difference in muscle activation patterns in the lumbar region between healthy and LBP groups [28], [100], like differences pertaining to FP and AEP patterns with motor control impairments considered to be the mechanism for maintaining CLBP. Also, altered temporal patterns of co-activation [98] and activity (re)distribution [52] between muscle sites in the lumbar region can be observed in LBP patients, compared to healthy subjects, implying different activity alignments for otherwise highly synergistic behavior of back muscles [39]. In that course, an interest to capture such temporal activation events on a finer time-localized scope was raised. Furthermore, such observations are inherently easier to capture for the functional type of exercises performed, due to distinctive time-related events (e.g. sitting vs. standing, flexion vs. extension, etc.) that are analyzed and compared in the course of the time. However, for isometric exercise, as in our case, a new measure (score) was required to capture different time-related events within and between muscle sites (Fig. 3), without an obvious external event-based trigger that would induce time-related myoelectric changes. For that reason, ''alignment'' and ''misalignment'' scores, for the respective primary feature in behind, were calculated. In this case, ''alignment'' assumed time-aligned co-activation (synergistic activation) across the majority of muscle sites, whereas for ''misalignment'', only individual muscle (single muscle site) activations were observed, i.e. no time-aligned co-activation across other muscle sites was observed.

To support further the construction of a new comprehensive model, fatigue indices and trends were included in the secondary features groups as well. Fatigue assessment, based on MDF linear regression, is a common concept utilized in the analysis of LBP, although not always straightforward results were reported. Namely, some papers reported higher values for the initial median frequency, f_0 (15) in LBP patients with statistical significance [25], whereas in some papers no significant difference between healthy and individuals with LBP was reported [13]. For the linear regression slope, k_{sl} (15), the reported results were also not univocally straightforward, however, most of the papers report statistically more significant results suggesting higher fatigability of muscles in the lumbar region for subjects with LBP [42], [75], [76]. To expand features set in a search of potentially useful LBP indices, beyond MDF and linear regression only, the Trends feature group was introduced as a set of measures setting the relations for both intra-channel and between channels (bilaterally and ipsilaterally), as presented in section IV-D.

B. HOW TO SELECT THE INITIAL FEATURE SET?

As the first step in the feature modeling process, fourteen primary features (Table 2) were chosen as a basis to find the best-performing ones in the context of LBP detection. The focus was on TD features as the most commonly employed feature group in sEMG-based classification [59], [61]. Alongside, some of the TD features were a common choice for many LBP-related studies, analyzing changes in EMG amplitude for activation detection or exhibited force levels via RMS, VAR, MAV or IEMG [6], [19], [20], [26]. Our initial set of primary (raw) features was further populated with KURT and SKEW, as higher-order statistics measures analyzed in [62], [104], MDF as used in a number of LBP related studies [20], [21], [25]–[27], [30], PE as a measure of non-linear signal complexity [64]. Finally, RVD, as a newly constructed feature, was added to the feature set with the intention to detect local changes in sEMG due to force level changes, activation bursts, or co-activation triggers. Such a feature was not found in any of the analyzed studies.

It was observed that the MDF feature demonstrated below-average individual classification results, which coincides with observations reported by Phinyomark *et al.* [59], stating that EMG features calculated in the frequency domain do not provide good results in EMG signal classification. However, interestingly, SSC feature, characterized by carrying certain frequency-related information, is shown to be one of the best features in sEMG based classifications, namely, being part of Hudgin's and Du's feature vectors, as well as the best ranking individual feature in this study.

Alongside, correlation analysis confirmed high correlation and redundancy among subset of features (MAV, IEMG, LD, RMS, VAR) (Fig. 6), where in our case best performing feature among these was VAR, unlike MAV in [57], [59], or RMS as employed in many studies [19], [20], [26]. It is also noteworthy to state that no bias towards boosting best feature-classifier pairs was promoted, but results across all classifiers were taken into account for the selection of primary features subset that was further employed in constructing secondary feature groups.

C. DO CONTEXTUAL FEATURES CONTRIBUTE TO LBP DETECTION?

We analyzed the classification accuracy results for each of the secondary feature groups individually (Table 5). It was observed that none of the feature groups excelled in detecting LBP patients alone. Best classification accuracy results, among created feature groups, were demonstrated for the coordination and co-activation measures with 0.91 (for fine kNN) and 0.88 (coarse SVM), respectively, where median accuracy results were 0.81 and 0.80, respectively. However, when all features groups were combined as input into the classifier models, the Complete set of features demonstrated improved results with higher accuracy (0.90, 0.91 for the fine

resolution and subspace *k*NN, respectively, and 0.93, 0.94, 0.94 for the medium Gaussian, cubic and quadratic kernel SVM classifiers, respectively), as well as an improved median accuracy (0.84) calculated across all classifiers.

These results contribute to the idea of introducing a multifactorial approach for describing LBP through feature modeling, as we hypothesized. Such concept was also following the questions and comments posed by Dieën *et al.* on differences in motor control between individuals with and without LBP [101] and whether such variations, in motor control issues, might contribute to more personalized approaches in LBP diagnostics and rehabilitation consequently [105], thus avoiding ''one-size fits all'' approach [5], [34].

Although features' diversity inherently assumes introducing more complexity into the classification models, this is inevitable if an individualized approach is aimed to be established. This was evidenced by the number of NCA components required to describe our complete model for detecting subjects with LBP (twenty components employed) compared to coordination and co-activation models alone (with eleven and twelve NCA components, respectively). However, by introducing contextual and more meaningful features and measures, the portion of complexity can be lifted through more immediate medical interpretation. This is contrary compared to some other studies where a significant number of features was employed, but with more difficult to interpret contributions of respective features [20], [39]. This challenge especially holds when deep learning techniques are employed [22]. Additionally, it is also worth acknowledging that an excessive number of final feature components, compared to the number of data samples, can lead to overfitting if the feature modeling procedure is not properly handled.

Moreover, different secondary feature subsets were observed across all cohorts' differentiation pairs as presented in Table 7, where Coordination, Co-activation, and Trends, as major feature groups, were present in all models. This suggests that diverse secondary feature groups might indeed support favorably LBP detection and differentiation through the proposed measures of contextual modeling. Otherwise, classification and differentiation results, based solely on the primary feature group and fatigue indices, with no constraint on the NCA components size, and submitted to the same feature selections procedure, would have demonstrated somewhat more success itself, as promoted in other studies, especially considering the relevant features pool given at hand (twenty-six different fatigue-based features and six primary features per each channel). In this study, that was shown not to be the case, where moreover, coordination and co-activation characteristics were the key LBP modeling contributors.

D. DO DIFFERENTIATION MODELS PROVIDE SATISFACTORY RESULTS?

The proposed method and the respective modeling demonstrated to be successful in differentiating the LBP group (a group combined from CLBP and RLBP patients) against HS, with the accuracy of up to 0.94 and high sensitivity of up to 0.97 (Table 6). However, for such accuracy results, the positive contribution of successfully detecting patients with RLBP is likely imminent. Namely, HS vs. RLBP classification pair provided median accuracy of 0.96, alongside high precision and sensitivity, with median values of 0.97 and 0.99 across all classifiers, respectively. Nonetheless, it is deemed fair to acknowledge that HS vs. LBP model had to account for a significant variation among groups and subjects at test, which also resulted in the largest number of NCA components (twenty), pointing to a complex model containing 4 out of 5 secondary feature groups (only MDF-based fatigue indices were not included in this particular model). On the other hand, differentiation between CLBP and HS groups, only, was somewhat less successful, but still with satisfactory results for the best classification results in the range of 0.88 to 0.90. Similarly, differentiation between CLBP and RLBP showed to be a challenging task, where fourteen NCA components were used to describe the model with at best classification accuracy of 0.86 and 0.89 achieved for subspace and fine *k*NN classifier, respectively. Furthermore, CLBP vs. RLBP model was the most diverse model including all of the secondary features group, thus pointing to motor control and neuromuscular complexity (or variability) in behind. Overall, these results are in alignment with so far commented challenges in defining classification models for non-specific CLBP problem description, as well as related to subgrouping tasks, as indicated in the case of CLBP vs. RLBP.

Furthermore, in all of the classification pairs, time-to-endurance (T_{EMG}), as a single non-sEMG parameter utilized in this study, did not contribute as a standalone parameter to any of the classification models. This parameter alone is generally shown to be a good indicator for the presence of LBP in different studies [6], [13], [75], although sometimes with contradictory results, as in work by Lariviere *et al.* where for healthy subjects shorter endurance was reported [106].

Additionally, it is also worth stating that the initial primary feature selection, resulting in SSC, PE, WAMP, VAR, ZC, and RVD feature set, was defined based on the best fit for HS vs. LBP classification. Moreover, all classifiers were employed out-of-the-box with a predefined set of parameters. This is leaving room for further parameter optimizations that could lead to the improved classification and differentiation results for more challenging tasks (namely HS vs. CLBP, or CLBP vs. RLBP).

When analyzing the classification results across twenty-three classifiers and their variants exploited, it is shown that SVM and *k*NN-based classifiers were consistently providing the best classification accuracy results. More concretely, quadratic and cubic kernel SVM, together with fine resolution and ensemble subspace *k*NN, demonstrated to be the most successful in dealing with LBP detection and related differentiation tasks. On the other hand, classifiers based on discriminant analysis and decision trees (together with ensemble boosted and bagged trees) were the least

successful. Below average classification success results for LDA were somewhat surprising as it was one of the most commonly employed classifiers in LBP classification related tasks [12], [24]–[27]. Somewhat better results were demonstrated by QDA, as dealing more successfully with more complex classification tasks in the quadratic separation surface, however, often suffering from resulting singular covariance matrix calculations.

In this study, a certain caution was posed around *k*NN-based classifiers. Namely, the fine resolution *k*NN classifier was labeling each sample at test based on only one ($k = 1$) nearest neighbor. Such an approach is very prone to outliers and noise. Additionally, NCA-based feature selection is based on a 1-NN classifier, thus certain bias towards respective *k*NN classification models might have been induced. To assess these effects on concrete LBP classification tasks, further investigation would be required. However, the obtained results in this study showed that SVM outperformed *k*NN classifiers, so it was deemed NCA not to have a decisive bias-prone impact on the classification success for the methods proposed. Moreover, for the ensemble subspace *k*NN models a medium resolution with $k = 10$ was employed. Furthermore, certain limitations might have been introduced with the feature selection procedure being performed on the whole data set. In an ideal case, the resulting feature models (i.e. feature selection) should have been validated on the data not being used for features (attribute) selection [97]. Otherwise, the information leakage and overfitting might have occurred, where for the filter-like selection approaches [107] the area under the receiver operating characteristic curve (ROC) might have grown for 5-10%, as for the scenarios demonstrated by Smialowski *et al.* [97].

E. CAN WE INTERPRET CLASSIFICATION MODELS?

Interpretability has been posed as another goal of this study, alongside satisfactory classification results. By employing the feature selection based on NCA, we have preserved the features in the original domain. This, combined with contextual feature modeling, has opened the opportunity to introduce more meaningful explanations in the classification models, giving more insights into the individualized sEMG-expressed characteristics related to LBP.

By taking the example of HS and RLBP differentiation, we observe that RLBP can be differentiated from HS with a very high accuracy (0.96 median value) by only taking three feature components (Table 7). Moreover, such a strong classification distinction can be translated into a discriminative visual representation, as shown in Fig. 8, with a very clear separation of grouping between the RLBP and HS. This was confirmed with the Wilcoxon rank sum test at 5% significance level. Furthermore, such visualization now provides a direct insight into the variations of the underlying RLBP characteristics (when compared with the HS group). It can be observed that these contextual characteristics are dominantly related to the LLES region with coordination (bilateral) and co-activation (synergistic time-alignment) aspects. Namely, the provided information is suggesting that patients with RLBP are dominantly exhibiting reduced bilateral coordination and co-activation in the LLES region compared to HS individually. Also, looking into the SSC-based feature (Tr-2) with a slower decreasing trend for RLBP patients, one of the interpretations could be that, in RLBP patients, one side of ULES is less engaged in muscle activity, thus less exposed to muscle fatigue. These interpretations follow some of the previous observations reported on asymmetrical functioning patterns in lower back muscles in patients having radiculopathy [30], where the reported asymmetry is likely to be related to the characteristic one-sided radiating pain.

On the other hand, differentiation between HS and CLBP introduced more components (thirteen) into the model, thus leading to an assumption of a more complex, i.e. more diverse, neuromuscular and motor control dynamics in behind. This complexity and diversity can be captured directly from Fig. 9 where two CLBP profiles for two different patients (left and right plot) are shown. Also, it can be observed that no clear separation between HS and CLBP groups is possible, moreover indicating quite heterogeneous CLBP profiles within the same group category (i.e. non-specific CLBP), unlike to more homogeneous RLBP case. However, this again concords with so far discussions on multifold variations in LBP patients and the necessity of subgrouping to tackle non-specific CLBP [34], [101]. If analyzing the CLBP patient's profile on the right side of Fig. 9, we observe that all coordination, co-activation, muscle activity, and fatiguability are significantly inhibited compared to median values for both the HS group and CLBP group. On the contrary, CLBP patient's profile on the left side demonstrated a significant muscle activity and changes in the LLES region, even exceeding statistically significant high values characteristic for the HS group (components 1, 5, 7, 9), alongside more muscle activity and co-activation present in ULES region (components 2 and 11) that are somewhat more characteristic to CLBP group itself. Furthermore, two additional expressions were observed. As the first one, a significant LLES L-R coordination for more complex muscle activity patterns in behind (component 7), and quite inhibited LLES L-R coordination for the exerted contraction force levels. For the HS group in general, these two coordination aspects are more balanced on the individual subject level, whereas this particular coordination discrepancy pattern has been observed in several other CLBP patients as well. As the second characteristic expression, on LLES left side, excessive muscle activity was detected, resulting in lots of hits and misses in co-activation triggering (components 9 and 10), where, overall, more synergistic misalignments were detected (components 10 and 12), compared to HS group.

One of the possible explanations for these two very distinctive profiles, observed across CLBP patients (Fig. 9), is in that patients adapt to pain due to musculoskeletal deficiencies in manifold ways. In some cases, these adaptations lead to avoidance-like inhibited behavior, or to an increased

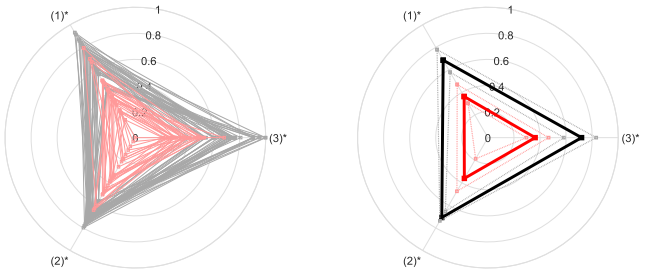

FIGURE 8. NCA component values are shown in spider plots for comparison between HS and RLBP subjects. NCA component values for each HS (gray) and RLBP (red) subject, are shown on left. Median values for each feature component, for HS and RLBP group, in black and red solid contours, shown on right, respectively. Dotted contours, inner and outer, for 25- and 75-percentile, shown for HS and RLBP group, in gray and light red, respectively. Three resulting NCA components, (1) to (3), as listed in Table 7 under HS vs. RLBP comparison. Components with * indicate the Wilcoxon rank sum test with a difference of median values at 5% significance level.

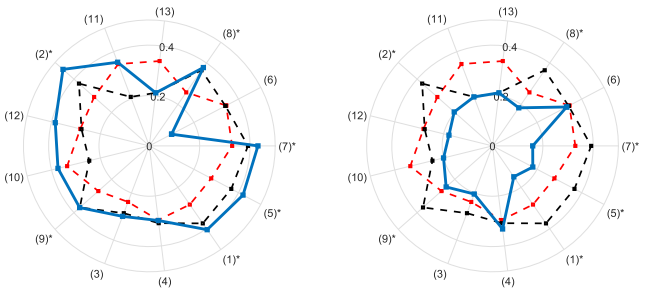

FIGURE 9. NCA component values for two individuals with CLBP are shown in spider plots with blue contours (left and right). Median component values across all HS and CLBP subjects are shown in black and red dashed contours, respectively. NCA components are labelled with (1) to (13) and referencing to Table 7. Components are organized in such a manner to follow the geometric placement of muscle sites, with components 8, 13, 11, and 2 relating to ULES (top) and the remaining components relating to LLES (bottom). The notion for left and right geometric placement is followed where possible. Components with * indicate the Wilcoxon rank sum test with a difference of median values at 5% significance level.

muscle activity to protect and compensate for the painful parts [37]. For the given profile examples, it is noteworthy to state that the endurance times (T_{EMG}) were 49 and 128 seconds, for the inhibited and increased muscle activity cases, respectively. For such outcomes, the psychological and anticipatory pain avoidance effects, reflected through the neuromuscular coding, also play a role.

Results, like those here presented, might further contribute to the insights into the inconsistencies reported in the number of studies that result in contradictory interpretations. Elaborated commentary on these challenges were also given by Dieën *et al.* [6], [101].

F. CONTRIBUTIONS

A novel approach for LBP patients detection and differentiation with contextual feature modeling has been proposed. Contextual LBP characteristics were modeled via secondary feature groups (Coordination, Co-activation, Trends, and Fatigue indices) derived from the simple raw (primary) features. NCA, as the feature selection method, was employed with the aim to preserve the features in the original domain. Therefore, a basis for further LBP patients' subgrouping has been established by providing a closer insight into the LBP-related neuromuscular and motor control characteristics. Additionally, new visualization of discriminating components via patient profiles has been proposed, thus facilitating medical interpretations.

The experiment for the proposed method had been arranged in a minimal setup to avoid manifold biases, where the primary focus was to validate the very concept of contextual feature modeling with corresponding expert interpretation. The frame for the setup was as follows: simple isometric trunk extension exercise, only two pairs of bilateral sEMG probes in the lumbar region (ULES and LLES), and ready classifiers with a predefined set of parameters, and with no additional fine-tuning.

The initial (primary) TD feature set, consisting of six features SSC, PE, WAMP, VAR, ZC, and RVD, has been proposed, validated, and consequently used to create secondary feature groups. A novel RVD feature had been created and introduced in this study to capture local myoelectric changes in muscle force or energy levels. By checking the NCA components in Table 7, it can be observed that the features based on RVD took a significant part in all of the resulting classification models, especially useful in co-activation triggers analysis. This suggests RVD to be a relevant feature choice and good discriminator in respective sEMG-based contextual LBP models. Similarly, PE is contributing in all models, except for HS vs. RLBP differentiation, for which a less complex model pertains.

Also, DTW and Spearman distance have been validated as similarity metrics for LBP coordination measures. These metrics have not been seen in other LBP studies, however, in this study, they have contributed to the final results (unlike some other metrics, like Euclidean distance or MI), thus proving a certain level of their usability in LBP analysis for more complex modeling, especially for the most challenging CLBP vs. RLBP differentiation tasks. In this study, we have also presented an approach for calculating coordination similarity measures intra-subject directly between muscle sites on time-aligned feature sequences accounting for local changes.

A novel contextual measure for tracking the co-activation alignments, across all muscle sites examined, has been constructed and employed. Co-activation scores have shown to be, alongside coordination-based features, the most contributing features to the overall classification results and models. Additionally, the benefit of applying autocorrelation to co-activation score measures has been confirmed. This approach proved to be useful in additionally improving the discriminating power of selected features, for all classification pairs except for HS vs. RLBP. One possible explanation for its success is that the autocorrelation by its nature has emphasized periodic or regular patterns of muscle activations, whereas, cases with irregular activation patterns

TABLE 7. NCA components lists and comparison as a result of different subject groups classification and differentiation for HS vs. LBP, HS vs. RLBP, HS vs. CLBP, CLBP vs. RLBP, from left to right, respectively. Secondary feature group (Group), feature metric (Metric), and involved relations among muscle locations (Relation), as feature insights, are listed for each classification group. Groups labeled CO, CA, TR, FI, and PR, stand for Coordination, Co-activation, Trends, Fatigue indices, and Primary-derived secondary feature groups, respectively. Each group's Metric provides information about the primary feature behind it (except for FI with implicit MDF). For the CO group, DTW, Corr, and Spear stand for dynamic time warping, Pearson correlation, and Spearman distance metrics, respectively. For the CA group, s- and u- stand for respective co-activation alignment ("sync") and misalignment ("unsync") scores. For TR group, Tr-1 to Tr-5 stand for features given with respective equations (10)-(14), with including respective muscle locations relations. For the FI group, FI-1 to FI-6 stand for fatigue indices as per section IV-E. The PR group is represented with the max value calculated on the respective raw (primary) feature. Muscle location relations are expressed in terms of different relation combinations for ULES (UL) and LLES (LL) lumbar regions, left (L) and right side, and starting (st.) or ending (end) segment of the exercise. The number of NCA components are twenty, three, thirteen, and fourteen for HS vs. LBP, HS vs. RLBP, HS vs. CLBP, and CLBP vs. RLBP, respectively.

NCA	HS vs. LBP			HS vs. RLBP			HS vs. CLBP			CLBP vs. RLBP		
	Group	Metric	Relation	Group	Metric	Relation	Group	Metric	Relation	Group	Metric	Relation
1	PR	max ZC	LL-R	TR	Tr-2 SSC	min UL	TR	Tr-1 SSC	min LL	PR	max ZC	LL-R
2	PR	max SSC	UL-L	CO	Corr SSC	LL L-R	TR	Tr-2 SSC	min UL	TR	Tr-4 VAR	LL L-R st.
3	PR	max VAR	LL-R	CA	s-RVD	LL-L	TR	Tr-4 WAMP	LL L-R end	TR	Tr-3 RVD	max LL-UL
4	TR	Tr-3 ZC	max LL-UL				TR	Tr-4 RVD	LL L-R end	TR	Tr-4 PE	LL L-R st.
5	TR	Tr-1 SSC	min LL				TR	Tr-1 PE	min LL	FI	FI-0	LL-R f0
6	TR	Tr-2 SSC	min UL				CO	Corr VAR	LL L-R	FI	FI-1	LL L-R k
7	TR	Tr-3 SSC	min LL-UL				CO	Corr PE	LL L-R	FI	FI-6	LL-R
8	TR	Tr-4 RVD	LL L-R st.				CA	s-SSC	UL-R	CO*	DTW SSC	LL L-R
9	TR	Tr-4 RVD	LL L-R end				CA	s-WAMP	LL-L	CO*	DTW RVD	UL L-R
10	CO	Corr VAR	LL L-R				CA	u-WAMP	LL-L	CO*	Spear PE	UL L-R
11	CO	Corr RVD	UL L-R				CA*	s-SSC	UL-L	CA	s-PE	LL-R
12	CA	s-VAR	LL-R				CA*	u-WAMP	LL-L	CA*	u-WAMP	LL-L
13	CA	u-VAR	UL-R				CA*	s-PE	UL-R	CA*	s-RVD	UL-L
14	CA	s-WAMP	LL-R							CA*	u-RVD	UL-R
15	CA	u-WAMP	UL-L									
16	CA	s-RVD	LL-L									
17	CA	s-RVD	UL-R									
18	CA*	s-WAMP	LL-L									
19	CA*	s-RVD	UL-R									
20	CA*	s-PE	UL-R									

* Groups with autocorrelation derived feature.

were diminished, thus further separating away those samples with initially less mutually distinctive co-activation triggering (as assumably more present in HS vs. CLBP case). In HS vs. RLBP case, this distinction was evident even without autocorrelation transformation, which seems likely possible due to significant motor control and myoelectric differences between HS and RLBP.

The performance of twenty-three classifiers and their variants have been validated and the best classifiers have been detected, namely, cubic and quadratic kernel SVM, fine resolution *k*NN, and ensemble subspace *k*NN classifier. Very good classification results have been achieved for different classification pairs between subject groups, especially for the LBP detection (CLBP and RLBP against HS) and RLBP differentiation (from HS), alongside maintaining meaningful medical interpretation.

G. FUTURE WORK

We presented this study and the respective methodology as part of a broader effort in building CDSS for supporting medical experts in subgrouping of LBP patients together with medical interpretation in behind. One of the main future efforts will be to establish not only qualitative but also quantitative-based profiling for individuals with LBP pathology. Thus, the intention is to more precisely measure inter-relations of different contributing factors (components or features) that result in potentially very distinctive individual LBP profiles, as shown in this study (Fig. 9). By categorizing more similar profiles, based on qualitative interpretation and quantitative measures, we believe additional advancements in inferring underlying neuromuscular and motor control processes could be made, hence leading to more personalized care with better confidence in the final rehabilitation outcomes. Additional insights might be drawn by correlating results with ODI, RDQ, and VAS indicators as well. Altogether, the resulting procedure should be maintainable in such a manner to enable the applicability of the respective method in clinical practice.

Another concurrent path, for further research, is to introduce additional effects into the proposed LBP contextual feature modeling, like analyzing contralateral and ipsilateral relations between lumbar regions muscle sites that might provide additional insights into the corresponding coordination aspects in LBP patients.

Future work further aims to tackle the verification of the proposed method on other unseen sets of subjects (with and without LBP) from different cohorts involved. This would reinforce the method's generalization and support removing any potential validation biases.

Finally, we believe that including additional muscle sites or bringing in more complex exercise tasks, would provide more insights into the underlying muscle activity mechanisms, although posing the questions on the complexity of the resulting models and subsequent interpretability of such

results. These aspects are worth exploring in the future with an aim to find a good balance among the setup complexity, amount of information and derived interpretation.

REFERENCES

[1] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, and A. Abdelalim, "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017," *Lancet*, vol. 392, pp. 1789–1858, Nov. 2018.

[2] J. L. Dieleman, R. Baral, M. Birger, A. L. Bui, A. Bulchis, A. Chapin, H. Hamavid, C. Horst, E. K. Johnson, and J. Joseph, "U.S. spending on personal health care and public health, 1996–2013," *J. Amer. Med. Assoc.*, vol. 316, no. 24, pp. 2627–2646, Dec. 2016.

[3] L. D. Bardin, P. King, and C. G. Maher, "Diagnostic triage for low back pain: A practical approach for primary care," *Med. J. Aust.*, vol. 206, no. 6, pp. 268–273, Apr. 2017.

[4] P. W. Hodges, "Pain and motor control: From the laboratory to rehabilitation," *J. Electromyogr. Kinesiol.*, vol. 21, no. 2, pp. 220–228, Apr. 2011.

[5] M. Rabey, A. Smith, P. Kent, D. Beales, H. Slater, and P. O'Sullivan, "Chronic low back pain is highly individualised: Patterns of classification across three unidimensional subgrouping analyses," *Scandin. J. Pain*, vol. 19, no. 4, pp. 743–753, Oct. 2019.

[6] J. H. van Dieën, L. P. J. Selen, and J. Cholewicki, "Trunk muscle activation in low-back pain patients, an analysis of the literature," *J. Electromyogr. Kinesiol.*, vol. 13, no. 4, pp. 333–351, Aug. 2003.

[7] B. Bazrgari and T. Xia, "Application of advanced biomechanical methods in studying low back pain–recent development in estimation of lower back loads and large-array surface electromyography and findings," *J. Pain Res.*, vol. 10, p. 1677, Jan. 2017.

[8] S. D. Tagliaferri, M. Angelova, X. Zhao, P. J. Owen, C. T. Miller, T. Wilkin, and D. L. Belavy, "Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: Three systematic reviews," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–16, Dec. 2020.

[9] L. Kalichman, D. H. Kim, L. Li, A. Guermazi, and D. J. Hunter, "Computed tomography–evaluated features of spinal degeneration: Prevalence, intercorrelation, and association with self-reported low back pain," *Spine J.*, vol. 10, no. 3, pp. 200–208, Mar. 2010.

[10] D. S. Teyhen, N. W. Gill, J. L. Whittaker, S. M. Henry, J. A. Hides, and P. Hodges, "Rehabilitative ultrasound imaging of the abdominal muscles," *J. Orthopaedic Sports Phys. Therapy*, vol. 37, no. 8, pp. 450–466, Aug. 2007.

[11] H. Ung, J. E. Brown, K. A. Johnson, J. Younger, J. Hush, and S. Mackey, "Multivariate classification of structural MRI data detects chronic low back pain," *Cerebral Cortex*, vol. 24, no. 4, pp. 1037–1044, Apr. 2014.

[12] S. H. Roy, C. J. De Luca, M. Emley, and R. J. C. Buijs, "Spectral electromyographic assessment of back muscles in patients with low back muscles in patients with low back pain undergoing rehabilitation," *Spine*, vol. 20, no. 1, pp. 38–48, Jan. 1995.

[13] M. Kankaanpää, S. Taimela, D. Laaksonen, O. Hänninen, and O. Airaksinen, "Back and hip extensor fatigability in chronic low back pain patients and controls," *Arch. Phys. Med. Rehabil.*, vol. 79, no. 4, pp. 412–417, Apr. 1998.

[14] M. E. Geisser, M. Ranavaya, A. J. Haig, R. S. Roth, R. Zucker, C. Ambroz, and M. Caruso, "A meta-analytic review of surface electromyography among persons with low back pain and normal, healthy controls," *J. Pain*, vol. 6, no. 11, pp. 711–726, Nov. 2005.

[15] B. U. Kleine, J. P. van Dijk, B. G. Lapatki, M. J. Zwarts, and D. F. Stegeman, "Using two-dimensional spatial information in decomposition of surface EMG signals," *J. Electromyogr. Kinesiol.*, vol. 17, no. 5, pp. 535–548, 2007.

[16] M. van der Hulst, M. M. Vollenbroek-Hutten, J. S. Rietman, and H. J. Hermens, "Lumbar and abdominal muscle activity during walking in subjects with chronic low back pain: Support of the 'guarding' hypothesis?" *J. Electromyogr. Kinesiol.*, vol. 20, no. 1, pp. 31–38, Feb. 2010.

[17] M. D. K. Jenssen, P. A. Bakkevoll, P. D. Ngo, A. Budrionis, A. J. Fagerlund, M. Tayefi, J. G. Bellika, and F. Godtliebsen, "Machine learning in chronic pain research: A scoping review," *Appl. Sci.*, vol. 11, no. 7, p. 3205, Apr. 2021.

[18] F. D'Antoni, F. Russo, L. Ambrosio, L. Bacco, L. Vollero, G. Vadalà, M. Merone, R. Papalia, and V. Denaro, "Artificial intelligence and computer aided diagnosis in chronic low back pain: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, p. 5971, May 2022.

[19] N. Jiang, K. D.-K. Luk, and Y. Hu, "A machine learning-based surface electromyography topography evaluation for prognostic prediction of functional restoration rehabilitation in chronic low back pain," *Spine*, vol. 42, no. 21, pp. 1635–1642, 2017.

[20] W. Du, O. M. Omisore, H. Li, K. Ivanov, S. Han, and L. Wang, "Recognition of chronic low back pain during lumbar spine movements based on surface electromyography signals," *IEEE Access*, vol. 6, pp. 65027–65042, 2018.

[21] M. Caza-Szoka, D. Massicotte, F. Nougarou, and M. Descarreaux, "Surrogate analysis of fractal dimensions from SEMG sensor array as a predictor of chronic low back pain," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 6409–6412.

[22] N. Wang, Z. Zhang, J. Xiao, and L. Cui, "DeepLap: A deep learning based non-specific low back pain symptomatic muscles recognition system," in *Proc. 16th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2019, pp. 1–9.

[23] B. X. W. Liew, D. Rugamer, A. M. D. Nunzio, and D. Falla, "Interpretable machine learning models for classifying low back pain status using functional physiological variables," *Eur. Spine J.*, vol. 29, no. 8, pp. 1845–1859, Aug. 2020.

[24] S. H. Roy, C. J. De Luca, M. Emley, L. I. Oddsson, R. J. Buijs, J. A. Levins, D. S. Newcombe, and J. F. Jabre, "Classification of back muscle impairment based on the surface electromyographic signal," *J. Rehabil. Res. Dev.*, vol. 34, no. 4, pp. 405–414, 1997.

[25] J. P. Peach and S. M. McGill, "Classification of low back pain with the use of spectral electromyogram parameters," *Spine*, vol. 23, no. 10, pp. 1117–1123, May 1998.

[26] S. I. Reger, A. Shah, T. C. Adams, J. Endredi, V. Ranganathan, G. H. Yue, V. Sahgal, and M. T. Finneran, "Classification of large array surface myoelectric potentials from subjects with and without low back pain," *J. Electromyogr. Kinesiol.*, vol. 16, no. 4, pp. 392–401, Aug. 2006.

[27] C. T. Candotti, J. F. Loss, A. M. S. Pressi, F. A. de Souza Castro, M. L. Torre, M. de Oliveira Melo, L. D. Araújo, and M. Pasini, "Electromyography for assessment of pain in low back muscles," *Phys. Therapy*, vol. 88, no. 9, pp. 1061–1067, Sep. 2008.

[28] W. Dankaerts, P. O'Sullivan, A. Burnett, L. Straker, P. Davey, and R. Gupta, "Discriminating healthy controls and two clinical subgroups of nonspecific chronic low back pain patients using trunk muscle activation and lumbosacral kinematics of postures and movements: A statistical classification model," *Spine*, vol. 34, no. 15, pp. 1610–1618, Jul. 2009.

[29] T. A. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. C. Williams, "Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 243–249.

[30] S. Ostojić, S. Peharec, V. Srhoj-Egekher, and M. Cifrek, "Differentiating patients with radiculopathy from chronic low back pain patients by single surface EMG parameter," *Automatika*, vol. 59, nos. 3–4, pp. 400–407, Oct. 2018.

[31] B. Hu, C. Kim, X. Ning, and X. Xu, "Using a deep learning network to recognise low back pain in static standing," *Ergonomics*, vol. 61, no. 10, pp. 1374–1381, Oct. 2018.

[32] A. Moniri, D. Terracina, J. Rodriguez-Manzano, P. H. Strutton, and P. Georgiou, "Real-time forecasting of sEMG features for trunk muscle fatigue using machine learning," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 718–727, Feb. 2021.

[33] F. Biering-Sørensen, "Physical measurements as risk indicators for low-back trouble over a one-year period," *Spine*, vol. 9, no. 2, pp. 106–119, Mar. 1984.

[34] N. E. Foster, J. C. Hill, and E. M. Hay, "Subgrouping patients with low back pain in primary care: Are we getting any better at it?" *Manual Therapy*, vol. 16, no. 1, pp. 3–8, Feb. 2011.

[35] D. Falla, V. Devecchi, D. Jiménez-Grande, D. Rúgamer, and B. X. W. Liew, "Machine learning approaches applied in spinal pain research," *J. Electromyogr. Kinesiol.*, vol. 61, Dec. 2021, Art. no. 102599.

[36] S. D. Tagliaferri, U. H. Mitchell, T. Saueressig, P. J. Owen, C. T. Miller, and D. L. Belavy, "Classification approaches for treating low back pain have small effects that are not clinically meaningful: A systematic review with meta-analysis," *J. Orthopaedic Sports Phys. Therapy*, vol. 52, no. 2, pp. 67–84, Feb. 2022.

[37] P. W. Hodges and K. Tucker, "Moving differently in pain: A new theory to explain the adaptation to pain," *Pain*, vol. 152, no. 3, pp. S90–S98, 2011.
[38] A. Jovic, I. Stancin, K. Friganovic, and M. Cifrek, "Clinical decision support systems in practice: Current status and challenges," in *Proc. 43rd Int. Conv. Inf., Commun. Electron. Technol. (MIPRO)*, Sep. 2020, pp. 355–360.
[39] C. Larivière, A. B. Arsenault, D. Gravel, D. Gagnon, and P. Loisel, "Evaluation of measurement strategies to increase the reliability of EMG indices to assess back muscle fatigue and recovery," *J. Electromyogr. Kinesiol.*, vol. 12, no. 2, pp. 91–102, Apr. 2002.
[40] A. Alexiev, "Some differences of the electromyographic erector spinae activity between normal subjects and low back pain patients during the generation of isometric trunk torque," *Electromyo. Clin. Neur.*, vol. 34, no. 8, pp. 495–499, 1994.
[41] N. Paquet, F. Malouin, and C. L. Richards, "Hip-spine movement interaction and muscle activation patterns during sagittal trunk movements in low back pain patients," *Spine*, vol. 19, no. 5, pp. 596–603, Mar. 1994.
[42] R. R. Hammill, J. R. Beazell, and J. M. Hart, "Neuromuscular consequences of low back pain and core dysfunction," *Clinics Sports Med.*, vol. 27, no. 3, pp. 449–462, Jul. 2008.
[43] B. Freriks and H. Hermens, "European recommendations for surface electromyography," *Roessingh Res. Develop.*, vol. 8, no. 2, pp. 13–54, 2000.
[44] R. Merletti and P. Di Torino, "Standards for reporting EMG data," *J. Electromyogr. Kinesiol.*, vol. 9, no. 1, pp. 3–4, 1999.
[45] R. Merletti and L. R. L. Conte, "Advances in processing of surface myoelectric signals: Part 1," *Med. Biol. Eng. Comput.*, vol. 33, no. 3, pp. 362–372, May 1995.
[46] T. W. Beck, T. J. Housh, G. O. Johnson, J. P. Weir, J. T. Cramer, J. W. Coburn, and M. H. Malek, "Comparison of Fourier and wavelet transform procedures for examining the mechanomyographic and electromyographic frequency domain responses during fatiguing isokinetic muscle actions of the biceps Brachii," *J. Electromyogr. Kinesiol.*, vol. 15, no. 2, pp. 190–199, Apr. 2005.
[47] V. Srhoj-Egekher, M. Cifrek, and V. Medved, "The application of Hilbert–Huang transform in the analysis of muscle fatigue during cyclic dynamic contractions," *Med. Biol. Eng. Comput.*, vol. 49, no. 6, pp. 659–669, Jun. 2011.
[48] H. Nieminen and E. Takala, "Evidence of deterministic chaos in the myoelectric signal," *Electromyo. Clin. Neur.*, vol. 36, no. 1, pp. 49–58, 1996.
[49] M. Lei, Z. Wang, and Z. Feng, "Detecting nonlinearity of action surface EMG signal," *Phys. Lett. A*, vol. 290, nos. 5–6, pp. 297–303, Nov. 2001.
[50] D. Farina, L. Fattorini, F. Felici, and G. Filligoi, "Nonlinear surface EMG analysis to detect changes of motor unit conduction velocity and synchronization," *J. Appl. Physiol.*, vol. 93, no. 5, pp. 1753–1763, Nov. 2002.
[51] D. Farina and R. Merletti, "Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions," *J. Electromyography Kinesiol.*, vol. 10, no. 5, pp. 337–349, 2000.
[52] Y. Hu, S. H. Siu, J. N. Mak, and K. D. Luk, "Lumbar muscle electromyographic dynamic topography during flexion-extension," *J. Electromyogr. Kinesiol.*, vol. 20, no. 2, pp. 246–255, Apr. 2010.
[53] M. Knaflitz and P. Bonato, "Time-frequency methods applied to muscle fatigue assessment during dynamic contractions," *J. Electromyogr. Kinesiol.*, vol. 9, no. 5, pp. 337–350, Oct. 1999.
[54] S. Karlsson, J. Yu, and M. Akay, "Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 2, pp. 228–238, Feb. 2000.
[55] D. T. Maclsaac, P. A. Parker, R. N. Scott, K. B. Englehart, and C. Duffley, "Influences of dynamic factors on myoelectric parameters," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 6, pp. 82–89, Nov. 2001.
[56] D. Farina, "Interpretation of the surface electromyogram in dynamic contractions," *Exerc. Sport Sci. Rev.*, vol. 34, no. 3, pp. 121–127, 2006.
[57] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
[58] Y.-C. Du, C.-H. Lin, L.-Y. Shyu, and T. Chen, "Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4283–4291, Jun. 2010.
[59] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, pp. 7420–7431, Jun. 2012.
[60] M. González-Izal, A. Malanda, E. Gorostiaga, and M. Izquierdo, "Electromyographic models to assess muscle fatigue," *J. Electromyogr. Kinesiol.*, vol. 22, no. 4, pp. 501–512, Aug. 2012.
[61] M. Hakonen, H. Piitulainen, and A. Visala, "Current state of digital signal processing in myoelectric interfaces and related applications," *Biomed. Signal Process. Control*, vol. 18, pp. 334–359, Apr. 2015.
[62] N. Nazmi, M. A. A. Rahman, S. Yamamoto, S. A. Ahmad, H. Zamzuri, and S. A. Mazlan, "A review of classification techniques of EMG signals during isotonic and isometric contractions," *Sensors*, vol. 16, no. 8, p. 1304, Aug. 2016.
[63] E. Olofsen, J. W. Sleigh, A. Dahan, and N. Zealand, "Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect," *Brit. J. Anaesthesia*, vol. 101, no. 6, pp. 810–821, Dec. 2008.
[64] X. Xi, M. Tang, S. M. Miran, and Z. Luo, "Evaluation of feature extraction and recognition for activity monitoring and fall detection based on wearable sEMG sensors," *Sensors*, vol. 17, no. 6, p. 1229, May 2017.
[65] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.
[66] V. Unakafova and K. Keller, "Efficiently measuring complexity on the basis of real-world data," *Entropy*, vol. 15, no. 12, pp. 4392–4415, Oct. 2013.
[67] K. W. N. Wong, J. C. Y. Leong, M.-K. Chan, K. D. K. Luk, and W. W. Lu, "The flexion–extension profile of lumbar spine in 100 healthy volunteers," *Spine*, vol. 29, no. 15, pp. 1636–1641, Aug. 2004.
[68] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.
[69] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2004, pp. 11–22.
[70] J. L. Myers, A. Well, and R. F. Lorch, *Research Design and Statistical Analysis*, 3rd ed. New York, NY, USA: Routledge, 2010.
[71] W. W. Daniel, *Applied Nonparametric Statistics*. Boston, MA, USA: Houghton Mifflin, 1978.
[72] H. M. Al-Angari, G. Kanitz, S. Tarantino, and C. Cipriani, "Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements," *Biomed. Signal Process. Control*, vol. 27, pp. 24–31, May 2016.
[73] X. Wu, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
[74] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić, "Surface EMG based muscle fatigue evaluation in biomechanics," *Clin. Biomech.*, vol. 24, no. 4, pp. 327–340, 2009.
[75] S. H. Roy, C. J. D. Luca, and D. A. Casavant, "Lumbar muscle fatigue and chronic lower back pain," *Spine*, vol. 14, no. 9, pp. 992–1001, Sep. 1989.
[76] A. F. Mannion, B. Connolly, K. Wood, and P. Dolan, "The use of surface EMG power spectral analysis in the evaluation of back muscle function," *J. Rehabil. Res. Dev.*, vol. 34, no. 4, pp. 427–439, 1997.
[77] J. D. Gibbons and S. Chakraborti, *Nonparametric Stat. Inference*, 6th ed. New York, NY, USA: Chapman & Hall, 2020.
[78] J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, L. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA, USA: MIT Press, 2004, pp. 513–520.
[79] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, pp. 161–168, Jan. 2012.
[80] K. Q. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Jul. 2009.
[81] *MATLAB Version 9.9.0.1524771 (R2020b) Update 2*, Mathworks, Inc., Natick, MA, USA, 2020.
[82] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification Regression Trees*. New York, NY, USA: Routledge, 1984.
[83] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
[84] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA, USA: Academic Press, 1990.

[85] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, vol. 35, nos. 5–6, pp. 352–359, 2002.
[86] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[87] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer, 2013.
[88] E. Fix and J. L. Hodges, Jr., "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *Int. Stat. Rev.*, vol. 57, no. 3, pp. 238–247, 1989.
[89] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, pp. 175–185, Aug. 1992.
[90] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[91] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, Jul. 1990.
[92] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[93] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors)," *Ann. Statist.*, vol. 28, no. 2, pp. 337–407, 2000.
[94] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
[95] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
[96] C. Ambroise and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 10, pp. 6562–6566, 2002.
[97] P. Smialowski, D. Frishman, and S. Kramer, "Pitfalls of supervised feature selection," *Bioinformatics*, vol. 26, no. 3, pp. 440–443, Feb. 2010.
[98] C. L. Hubley-Kozey and M. J. Vezina, "Differentiating temporal electromyographic waveforms between those with chronic low back pain and healthy controls," *Clin. Biomech.*, vol. 17, nos. 9–10, pp. 621–629, Nov. 2002.
[99] L. I. E. Oddsson and C. J. D. Luca, "Activation imbalances in lumbar spine muscles in the presence of chronic low back pain," *J. Appl. Physiol.*, vol. 94, no. 4, pp. 1410–1420, Apr. 2003.
[100] P. O'Sullivan, "Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impairments as underlying mechanism," *Manual Therapy*, vol. 10, no. 4, pp. 242–255, Nov. 2005.
[101] J. H. van Dieën, N. P. Reeves, G. Kawchuk, L. R. van Dillen, and P. W. Hodges, "Motor control changes in low back pain: Divergence in presentations and mechanisms," *J. Orthopaedic Sports Phys. Therapy*, vol. 49, no. 6, pp. 370–379, Jun. 2019.
[102] W. W. Lu, K. D. K. Luk, K. M. C. Cheung, Y. W. Wong, and J. C. Y. Leong, "Back muscle contraction patterns of patients with low back pain before and after rehabilitation treatment: An electromyographic evaluation," *J. Spinal Disorders*, vol. 14, no. 4, pp. 277–282, Aug. 2001.
[103] N. P. Reeves, J. Cholewicki, and S. P. Silfies, "Muscle activation imbalance and low-back injury in varsity athletes," *J. Electromyogr. Kinesiol.*, vol. 16, no. 3, pp. 264–272, Jun. 2006.
[104] R. H. Chowdhury, M. B. I. Reaz, M. A. B. M. Ali, A. A. A. Bakar, K. Chellappan, and T. G. Chang, "Surface electromyography signal processing and classification techniques," *Sensors*, vol. 13, no. 9, pp. 12431–12466, 2013.
[105] J. H. van Dieën, N. P. Reeves, G. Kawchuk, L. R. van Dillen, and P. W. Hodges, "Analysis of motor control in patients with low back pain: A key to personalized care?" *J. Orthopaedic Sports Phys. Therapy*, vol. 49, no. 6, pp. 380–388, Jun. 2019.
[106] C. Larivière, A. B. Arsenault, D. Gravel, D. Gagnon, and P. Loisel, "Surface electromyography assessment of back muscle intrinsic properties," *J. Electromyogr. Kinesiol.*, vol. 13, no. 4, pp. 305–318, Aug. 2003.
[107] R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Inform.*, vol. 85, pp. 189–203, Sep. 2018.

VEDRAN SRHOJ-EGEKHER (Student Member, IEEE) was born in Dubrovnik, Croatia, in 1984. He received the M.S. degree in electrical engineering from the University of Zagreb, Zagreb, in 2009, where he is currently pursuing the Ph.D. degree in electrical engineering. His graduate thesis was awarded as the Best Graduate Thesis in the field of biomedical engineering by the Croatian Medical and Biological Engineering Society (CROMBES) in 2010. He is the first author or coauthor of several conference and scientific papers with key interests in biomedical engineering, signal processing, machine learning, and medical image analysis.

MARIO CIFREK (Senior Member, IEEE) received the Dipl.-Ing., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 1987, 1992, and 1997, respectively. He is currently a Professor of electrical engineering at the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb. His research interests include biomedical instrumentation and biomedical signal analysis for research and clinical applications. He is a Senior Member of the IEEE Engineering in Medicine and Biology Society (EMB18 Chapter Chair from 2014 to 2017), the IEEE Instrumentation and Measurement Society, the IEEE Signal Processing Society, International Federation for Medical and Biological Engineering, the Croatian Biomedical Engineering and Medical Physics Society, and the Croatian Society for Communications Computing, Electronics, Measurement, and Control. He is a Full Member of the Croatian Academy of Engineering, where he has been the Secretary of the Department of Systems and Cybernetics, since 2017.

STANISLAV PEHAREC received the Dipl. Defectologist-Somatopedist degree from the Faculty of Defectology, University of Belgrade, in 1987, the M.Sc. degree from the Faculty of Kinesiology, University of Zagreb, in 2000, and the Ph.D. degree from the Faculty of Medicine, University of Rijeka, Croatia, in 2007. He is currently an Assistant Professor of biomechanics, diagnostics in physiotherapy and anatomy at the Faculty of Health Sciences, University of Rijeka; the Faculty of Medicine, and Diagnostics in Kinesiotherapy, Juraj Dobrila University of Pula; and the Faculty of Kinesiology, University of Zagreb. He is the author or coauthor of several conference and scientific papers focusing on biomechanics, physiotherapy, rehabilitation, and kinesiology. He is a member of the Croatian Vertebrological Society.