



Understanding Violin Players' Skill Level Based on Motion Capture: a Data-Driven Perspective

Vincenzo D'Amato¹ · Erica Volta¹ · Luca Oneto¹ · Gualtiero Volpe¹ · Antonio Camurri¹ · Davide Anguita¹

Received: 29 June 2020 / Accepted: 9 September 2020 / Published online: 24 September 2020
© The Author(s) 2020

Abstract

Learning to play and perform a music instrument is a complex cognitive task, requiring high conscious control and coordination of an impressive number of cognitive and sensorimotor skills. For professional violinists, there exists a physical connection with the instrument allowing the player to continuously manage the sound through sophisticated bowing techniques and fine hand movements. Hence, it is not surprising that great importance in violin training is given to right hand techniques, responsible for most of the sound produced. In this paper, our aim is to understand which motion features can be used to efficiently and effectively distinguish a professional performance from that of a student without exploiting sound-based features. We collected and made freely available a dataset consisting of motion capture recordings of different violinists with different skills performing different exercises covering different pedagogical and technical aspects. We then engineered peculiar features and trained a data-driven classifier to distinguish among two different levels of violinist experience, namely beginners and experts. In accordance with the hierarchy present in the dataset, we study two different scenarios: extrapolation with respect to different exercises and violinists. Furthermore, we study which features are the most predictive ones of the quality of a violinist to corroborate the significance of the results. The results, both in terms of accuracy and insight on the cognitive problem, support the proposal and support the use of the proposed technique as a support tool for students to monitor and enhance their home study and practice.

Keywords Movement technology · Music education · Music learning technology · Multimodal interactive systems · Machine learning · Feature engineering · Feature ranking · Model selection · Error estimation

Introduction

In music education, individual music practice is an essential element to teach from the first period of study and that accompanies the music student and then the concert musician throughout his musical career. In [1], researchers indicate that expert music performers based their work on a methodical rehearsals' planning and a systematic approach to problem identification, strategy selection, and evaluation, namely all components of self-regulated thinking. Consequently, in music education, the same approach is used. Self-regulated learning is a cognitive model that bases its efficiency on an effective planning and execution, and reviewed strategies to enhance learning. Moreover, this model also includes meta-strategies that require knowledge of the nature and benefit of each component of the learning process, and the ability to understand when to use it. These strategies include but are not limited to planning goals, mental rehearsal, and error understandings. Practice is a core element of music

✉ Luca Oneto
Luca.Oneto@unige.it

Vincenzo D'Amato
Vincenzo.DAmato@unige.it

Erica Volta
Erica.Volta@unige.it

Gualtiero Volpe
Gualtiero.Volpe@unige.it

Antonio Camurri
Antonio.Camurri@unige.it

Davide Anguita
Davide.Anguita@unige.it

¹ DIBRIS Department, University of Genoa, Via Opera Pia 13, I-16145, Genoa, Italy

education and career, either address individual or group and it represents the way musicians learn outside the classroom. It can be often believed that from a music lesson it is possible to infer enough information to successfully understand and face the issues that rehearsals may rise. In literature, however, it can be easily found indication that beginner music students do not demonstrate any systematic plan during their individual practice sessions [2, 3]. Hence, often, no self-learning occurs, inhibiting the musical development and progress of students [4]. Furthermore, music students learn how to play music in specific contexts (e.g., music classroom) in which they have a vast set of resources to use, but at the same time, they also face challenges in communication with teachers and a highly competitive and stressful environment where to develop their studies and careers. Moreover, the traditional education model is resistant to change and difficult to adapt to students' needs, for example one challenge related to traditional pedagogical method is that students have to master the difficulties to manage a coherent perspective between their proper proprioceptive feedback and teachers' suggestions. Traditional music teaching, indeed, is mostly based on a dyadic relationship between teacher and student, in which the time lag between students' performance and teachers' feedback makes the second to be dissociated from the auditory perception of the student, as it can be read in [5]. Since, often, the time in which this relation occurs is limited to weekly lectures, this element is even more important [6]. The long period of self-studying of students can be harsh and lead to a solitary process that often can also cause a high rate of abandonment [7]. In order to best address these challenges of music learning, it is particularly useful to reflect on reflective thinking and the cognitive dimension of learning. In [8], the author analyzed "how we think," distinguishing four forms of thinking that are naturally present in the human mind. The fourth type highlighted was identified as reflective thinking, which became a pillar of what we actually indicated as part of metacognition, the self-regulated thinking [9]. Using the words of [8]:

Reflection involves not simply a sequence of ideas, but a consequence – a consecutive ordering in such a way that each determines the next as its proper outcome, while each outcome in turn leans back on, or refers to, its predecessors. (Dewey 1933, p. 4)

Metacognition is generally understood as "thinking about one's own thoughts" [10], involving "active monitoring and consequent regulation and orchestration" [11] to complete a task. So, metacognition as part of the learning process can help identify what students are thinking while they engage in learning a task. In music educational curricula, traditionally, only overt behaviors were outlined as a means

of assessment, e.g., the final performance of a given repertoire. The evaluation in this way is based on students' ability to perform correctly the repertoire, taking care of pitches, rhythms, articulations, dynamics, accentuation, expressive phrasing, and so on. Rarely, the student is asked to demonstrate her/his musical understanding of music, or how s/he came to be able to perform a given repertoire. Nevertheless, it is important for students to be aware of what they know and self-assess their learning process. This important aspect of music teaching is often neglected and students are used to practice something on their own without understanding how to organize and evaluate their progress while rehearsing alone [12]. Research demonstrated that when students have the change to foster metacognitive skills, learning outcomes improve [12, 13]. In music education research, self-regulation strategies have been investigated as well, given new insights on how students learn and master music materials. Nielsen [14] studied how two college musicians used self-regulation strategies to monitor their learning outcomes. The author used observation of practice behaviors, verbal reports during practice session, and retrospective debriefing reports after the practice session to analysis of self-learning. With the same purpose of studying practice habits, Hallam [1] collected interviews of master musicians and beginners. The novice ones were also recorded playing a new piece after 10 min of practice session. Both these authors in their researchers identified students' application of self-learning strategies during their practice sessions. Also, the frequent use of repetition of segments or single notes during practice was reported. In [2], the authors examined practice strategies used by beginner instrumental students, from 7 to 9 years old, over a 3-year period. The studies here briefly presented provide useful insights into self-learning in music practice. Nevertheless, one more aspect has to be taken into account, investigating if and how music technology can be efficiently applied to educational settings, meaning the biomechanical aspect of music playing.

The biomechanical skills necessary for an accurate and safety performance, indeed, are often limited to subjective and vague perception and based on oral content transmission between the teacher and the students [15]. It seems, then, to be reasonable to suppose that more quantitative methodologies already tested and useful in other contexts, such as in sport medicine, could be applied and be useful also in understanding and teaching biomechanical skills of music performance [16].

In particular, the background knowledge that can be helpful in understanding motor skills needed in music performance, includes the motor learning theory and technology-based systems for analysis, monitoring, and evaluating learning efficiency [17]. In motor learning theories [18], three elements are presented as essential to

success: the characterization of the skills to be acquired, skills transfer between dissimilar systems, and skills acquisition without injuries. For an acquisition of the characterizing skills is necessary a scientific analysis to identify motor patterns, such as the coordination of neural and musculoskeletal systems. Moreover, the motor behaviors of professional players can be used as a reference model to facilitate the understanding of the essential skills to be transferred to music students. By directing attention to specific motor behaviors, learners can assimilate in an efficient and effective way the skills adequate to their technique. Following these findings, an emerging literature is more and more interested in investigating how full-body and motion analysis technologies may enhance music performance and learning outcomes, minimizing at the same time the risk of injuries [19–21].

An emerging literature is focusing on how movement analysis technologies can be used to inform music performance by enhancing learning outcomes and preventing risk of injuries [19, 22]. People are often prone to making mistakes during analysis or, possibly, when trying to establish relationships between multiple features. Machine learning (ML) can often be successfully applied to these problems to enhance these cognitive processes.

In this study, we present a system able to perform automated classification of highly professional musicians' and students' performance recordings, based on motion features analysis of selected violin techniques and repertoire. Our aim is to understand which motion features can be exploited to efficiently distinguish a professional performance and to use such information for real-time student assistive technologies. In literature, many ML models are able to classify information from time series and, in particular, from motion capture (MOCAP) data. These data provide a representation of the complex spatio-temporal structure of human motion. During a traditional MOCAP session, the locations of characteristic parts on the human body such as joints are recorded over time using appropriate devices. In the literature, it is possible to find several ways to handle MOCAP data. In particular, it is possible to identify two families of approaches [23, 24]. The first one, comprising traditional ML methods, needs an initial phase where the features must be manually and carefully extracted from the data [25–29]. The second family, which includes deep learning methods, automatically learns both features and models from the data [30, 31]. MOCAP data studies use both families of methods based on the cardinality of the sample size. For small cardinality dataset, deep learning-based methods cannot be employed since they require a huge amount of data to be reliable and to outperform traditional ML models with context-specific experience-based engineered features. For this reason, in most studies, traditional ML techniques are employed. These techniques have

already shown in the past to be successfully applied in the field of cognitive computation, in many applications [32–35]: from sequential learning [36] to sentiment analysis, as well as data management [37] and classification [38, 39].

In this work, we exploited the recording of four internationally renowned violin performers, selected by the Royal College of Music of London, and three novice students. The music selected encompassed 41 exercises, chosen by teachers of the Royal College of Music of London, from various sources of classical violin pedagogy literature. The exercises focused on several techniques, typical of a traditional pedagogical violin program, such as: handling the instrument, technique of the right and left hands, articulations studies, and repertoire pieces. Then, we propose the combination of carefully crafted features in combination with random forest (RF) to distinguish between the two skill levels of the violinist. RF [40] is considered a state-of-the-art learning algorithm for classification purposes since it has shown to be one of the most effective tools in this context [41, 42]. From a cognitive point of view, RF implements the wisdom of crowds principle, namely the aggregation of information in groups, resulting in decisions that are often better than could have been made by any single member of the group [43–45]. The main requirement behind this principle, which yields better results, is that there should be significant differences or diversity among the models. Many examples of the use of this principle in cognitive computation exist in the literature [44–50]. In accordance with the intrinsic hierarchy present in the data set, we will study two different scenarios: extrapolation with respect to different exercises and violinists. Furthermore, we will study which features are the most predictive ones of the quality of a violinist to corroborate the significance of the results. Results, both in terms of accuracy and insight on the problem, will support the proposal and the use of the proposed technique as a support tool for students to monitor and enhance their home study and practice.

The rest of the paper is organized as follows. The description of the problem and related data is reported in “[Data Description](#)”. The description of the proposed data-driven methodology is presented in “[Methods](#).” The results of applying the methods presented in “[Methods](#)” on the problem described in “[Data Description](#)” are reported in “[Experimental Results](#)”. Finally, “[Conclusions](#)” concludes the paper.

Data Description

The data employed in the present work were collected during H2020 ICT-TELMi Project¹. The project studied how

¹<http://telmi.upf.edu/>

Fig. 1 The violinist MOCAP setup: markers are placed over subject, bow, and violin



we learn violin playing and how technology should be designed to effectively support and enhance music instrument learning outcomes. For this reason, the collected corpus of data was designed as a collection of exercises to follow the learning path of classical violin conservatoire programs. It included several sources of data: MOCAP of the performer, violin, and bow (see Fig. 1), ambient and instrument audio and video, physiological data (electromyography) captured with Myo sensor², and Kinect data. The recordings took place at the Casa Paganini - InfoMus research center of the University of Genova³.

The recorded material was post-processed and uploaded into the repoVizz repository and made publicly available⁴. For a comprehensive description of the entire TELMI multimodal archive, refer to [51].

As previously mentioned, we recorded four internationally famous soloists, selected by the Royal College of Music of London, and three beginner violinists. For what concerns the chosen literature, modern violin students have hundreds of years of pedagogical material at their disposal, much of which is freely available online, via school and public libraries, and through their teachers. This explains the reason why the music selected for the TELMI project encompassed different exercises, chosen by teachers of the Royal College of Music of London, from various sources of classical violin pedagogy literature. The exercises focused on different techniques concerning the following: handling the instrument, technique of the right and left hands, articulations study, and repertoire pieces. The exercises take three forms: (1) those sourced from the standard published catalog of exercises, including those of Schradieck, Ševčík, and Kreutzer, which the survey the TELMI consortium did during the project found to be the most popular; (2) those sourced or adapted from the Associated Board of the Royal Schools of Music (ABRSM) examination syllabus; and (3)

customized exercises developed by Madeleine Mitchell to address specific techniques with specific focus on the capabilities offered by non-notated feedback (e.g., the bowing exercises). The use of both custom and pre-existing exercises was deliberate. Due to the national and international popularity of the ABRSM system, a subset of TELMI exercises was drawn from the scales and exercises for Grades 6 to 8 to represent the intermediate level of technical development and allow students to prepare for exams in which they may already be involved. Intonation, or tuning of the individual notes, is a core technique of violin playing. This skill will be tested in every exercise that involves a notated score, and especially in the scales for the beginning and intermediate students. As a supplement to this material, one of the most popular exercises comes from Schradieck's *School of Violin Technics* (1899), and in particular the first exercise from *Book I: Exercises for Promoting Dexterity in the Various Positions*. This exercise comprises a series of repeating scale patterns designed for careful and even control of note production. For shifting of the left hand between the seven positions on the neck of the violin, Yost's *Exercises for Changes of Position* (1928) was chosen. This collection systematically tests changes of every interval on every possible shift between the seven positions (1st to 2nd, 1, etc.) on each of the four strings. The transition from one string to the next requires a rotational motion of the bow in the right hand and shift of the fingers in the left. A classic exercise to develop this technique is Kreutzer's *String Crossing* from *Etude No. 13 for Solo Violin*. Kreutzer's *Etude No. 14 for Solo Violin* was chosen as a representative of the trill technique exercise. For the articulation exercises, the TELMI repertoire list proposed *Martelé* from Kreutzer's *Etude No. 7* Sautillé technique is emphasized in several of Ševčík's *Violin Studies, Op. 3*. For TELMI, Variation No. 16 was chosen. Spiccato articulation is again a technique emphasized in several of Ševčík's *Violin Studies, Op. 3*. For TELMI, Variation No. 34 was chosen. For the Staccato was selected Kreutzer, *Etude N. 4*. Finally, for the Arpeggios Flesch's *System of Scales* was chosen and as repertoire pieces one piece from Romantic literature and one

²<https://support.getmyo.com/hc/en-us/articles/202648103-Myo-Gesture-Control-Armband-tech-specs>

³www.casapaganini.org

⁴<http://telmi.upf.edu/opendatabase/>

Table 1 Description of the available dataset

	Violinists	Exercises	
Experts	4	11	(3 scales, 4 repertoires, and 4 techniques)
Beginners	3	8	(3 scales, 3 repertoires, and 2 techniques)
Total	7	19	(6 scales, 7 repertoires, and 6 techniques)

from a contemporary composer were chosen, namely they were Elgar's (1889) *Salut d'Amour, Op. 12*, and Michael Nymann's (2007) *Taking it as Read*.

From this archive of data, we selected one scale, one study, and one repertoire piece, namely a ABRSM Scale in G Major played *detaché* on 3 octaves, *String Crossing* from Kreutzer *Violin Study op.13, Salut d'Amour, Op.12* by Edgar, to start to investigate automatic classification from motion capture data. On this data, we computed 14 low-level features using the EyesWeb XMI platform⁵ [52]. The computed features are as follows: mean shoulders' velocity, shoulder low back asymmetry, upper body kinetic energy, left/right, bow–violin incidence, distance low/middle/upper bow–violin, hand–violin left/right head inclination, and left/right wrist roundness [21]. These features fully describe, based on the knowledge of expert players and teachers, the movements of the violinists. The final dataset included the 14 raw features described above of the 3 exercises played by each of the violinists. In particular, 7 violinists are recorded, of which 4 are experts and 3 are beginners. Nineteen files were made available containing information on the exercise performed by each musician (one exercise per class was missing). The dataset is summarized in Table 1.

In order to avoid noise due to the initial and final moments of the recording where musicians do not play, only timestamps where a piece of music is played are chosen.

Methods

The problem described in the previous sections can be easily mapped into a binary classification problem [23]. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be the input space (namely the features engineered from the different measurements), consisting of d features (see “[Features Engineering](#)”), and let $\mathcal{Y} = \{0, 1\}$ (namely expert and beginner) be the output space. Let $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, where $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y} \forall i \in \{1, \dots, n\}$, be a sequence of $n \in \mathbb{N}^*$ samples drawn from $\mathcal{X} \times \mathcal{Y}$. Let us consider a model (function) $f : \mathcal{X} \rightarrow \mathcal{Y}$ chosen from a set \mathcal{F} of possible hypotheses. An algorithm $\mathcal{A}_{\mathcal{H}} : \mathcal{D}_n \times \mathcal{F} \rightarrow \mathcal{F}$ characterized by its hyperparameters \mathcal{H} selects a model inside a set of possible ones based on the available dataset.

⁵http://www.infomus.org/eyesweb_eng.php

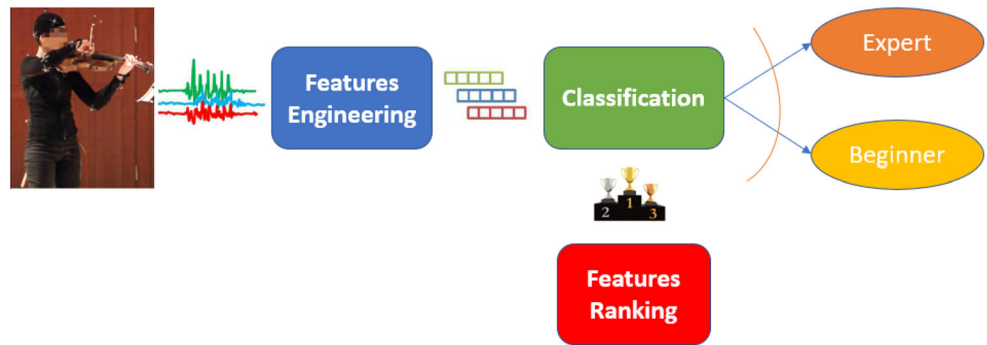
Note that many algorithms for solving binary classification problems exist in literature [23] but random forest (RF) has shown to be one of the most powerful ones [40–42] (see “[Random Forests](#)” and [Appendix A](#)). The error of f in approximating $\mathbb{P}\{Y | X\}$ is measured by a prescribed metric $M : \mathcal{F} \rightarrow \mathbb{R}$. Note also that many different metrics are available in literature for binary classification which may provide insights on the performance of the model [53] but the confusion matrix together with the accuracy (since in our case classes are balanced) is surely the most informative ones. In order to tune the performance of the $\mathcal{A}_{\mathcal{H}}$, namely to select the best set of hyperparameters, and to estimate the performance of the final model according to the desired metrics, model selection (MS) and error estimation (EE) phases need to be performed [54] (see “[Model Selection and Error Estimation](#)”). Moreover, in order to understand, from a cognitive point of view, how the algorithm exploits the derived features to make a prediction, a feature ranking phase is also performed (see “[Feature Ranking](#)”). The entire violinist skill-level classification process pipeline is as shown in [Fig. 2](#).

Features Engineering

In this section, we describe how the features have been extracted and engineered from the raw data described in “[Data Description](#)”.

The time series of the different measurements (left head inclination, left wrist roundness, etc.) were sampled in fixed-width sliding windows of 10 s. One second in the original time series acquired from MOCAP contains 1000 rows. Then, the fixed-width sliding windows contain information about 1000 rows times 10 s for a total of 10,000 rows, which is sufficient to capture cycles in human activities. Note that the fixed-width sliding windows have 50% of overlap in time. This heuristic has been already successfully employed in many different works in literature [55–57]. In order to analyze clear data and avoid noise due to the initial and final moments of recording, only timestamps where a piece of music is played are chosen. From each sampled window, a vector of features was obtained by computing standard measures previously employed in literature to describe human actions [57–60] such as the mean, the signal-pair correlation, and the signal magnitude area for both the time and frequency

Fig. 2 The violinist skill-level classification pipeline



domains (see Table 2). The fast Fourier transform was employed to find the frequency components for each window. A new set of features was also employed in order to improve the learning performance, including energy of different frequency bands, frequency skewness, and frequency kurtosis. Table 2 contains the list of all the measures applied to the time and frequency domain signals. This results in a total of 332 features. The resulting dataset has been made freely available for the research community⁶.

Random Forests

A powerful algorithm, both in terms of theoretical properties and practical effectiveness [41, 42], for classification is RF developed in [40] for the first time. In order to be able to fully understand RF, we need to recall how a binary decision tree (DT) [61] is defined and constructed. A binary DT for classification is a recursive binary three-structure in which a node represents a check on a particular feature, each branch defines the outcome of the check, and the leaf nodes represent the final classification. A particular path of exploration from the root of the tree to one of its leaves represents a classification rule. Based on a recursive schema, a DT is grown until it reaches a desired depth n_d . Each node of the DT (both root and nodes) is constructed by choosing the features and the check that most effectively separates the data satisfying the partial rule into two subsets based on the information gain (or possible other metrics like the classification accuracy). Given this definition of DT, it is then possible to understand RF and the learning phase of each of the n_t DT which composes the forest. From \mathcal{D}_n , a bootstrap sample (sample with replacement) \mathcal{D}' of n_b is extracted. Then, a DT is learned based on \mathcal{D}' , but the best check/cut is selected among a subset of n_v features over the possible n_f features randomly chosen at each node. n_d is set to infinite, namely the DT is grown until every sample of \mathcal{D}' is correctly classified. In the forward phase, namely the phase in which a previously unseen X needs to be labeled,

each DT composing the RF is exploited to classify X ; the final classification is taken with majority vote. Note that n_b, n_v, n_d , and n_t are the hyperparameters of the RF. If $n_b = n, n_v = \sqrt{d}$, and $n_d = \infty$, we obtain the original RF formulation [40], where n_t is usually chosen to trade-off accuracy and efficiency [62] since the larger it is the better.

The Metrics

For what concerns the metrics $M(f)$ exploited for evaluating the performance of a model f learned from the data based on the methods described above, we have to recall that many different metrics are available in literature [53]. In this work, we will report just the most common ones. In order to define them, let us first consider a subset of the available data \mathcal{T}_t , also called test set, coming from μ but different from \mathcal{D}_n since the data that have been used to learn f should be different from the ones exploited

Table 2 List of measures for computing feature vectors

Function	Description
mean	Mean value
var	Variance
mad	Median absolute value
max	Largest value in array
min	Smallest value in array
sma	Signal magnitude area
energy	Average sum of squares
iqr	Interquartile range
entropy	Signal entropy
correlation	Correlation coefficient
kurtosis	Signal kurtosis
skewness	Signal skewness
maxFreqInd	Largest frequency component
argMaxFreqInd	Index largest frequency component
meanFreq	Frequency signal weighted average
skewnessFreq	Frequency signal skewness
kurtosisFreq	Frequency signal kurtosis
ampSprec	Amplitude spectrum of the frequency signal
angle	Phase angle of the frequency signal

⁶<https://www.dropbox.com/sh/uwdvihjfy0uw13/AAD-IXsn-HsH5J0PZ9dGgtK4a>

to evaluate its performance so to avoid overfitting [54]. Let us define the element in the confusion matrix, the true positive ($TP(f) = \sum_{(X,Y) \in \mathcal{T}_m: Y=1} \mathbb{1}\{f(X) = 1\}$), the true negative ($TN(f) = \sum_{(X,Y) \in \mathcal{T}_m: Y=0} \mathbb{1}\{f(X) = 0\}$), the false positive ($FP(f) = \sum_{(X,Y) \in \mathcal{T}_m: Y=0} \mathbb{1}\{f(X) = 1\}$), and the false negative ($FN(f) = \sum_{(X,Y) \in \mathcal{T}_m: Y=1} \mathbb{1}\{f(X) = 0\}$), on this data. Then, we can also define the accuracy as:

$$\text{accuracy}(f) = \frac{TP(f) + TN(f)}{TP(f) + FN(f) + TN(f) + FP(f)}, \quad (1)$$

the precision as

$$\text{precision}(f) = \frac{TP(f)}{TP(f) + FP(f)}, \quad (2)$$

the recall as

$$\text{recall}(f) = \frac{TP(f)}{TP(f) + FN(f)}, \quad (3)$$

and the area under the receiver operating characteristic curve (ROC-AUC), which is the area under the $TP(f)$ rate against the $FP(f)$ rate curve.

Model Selection and Error Estimation

MS and EE face and address the problem of tuning and assessing the performance of a learning algorithm [54]. In this work, we will exploit the resampling techniques which leverage on a simple idea: \mathcal{D}_n is resampled many (n_r) times, with or without replacement, and three independent datasets called learning, validation, and test sets, respectively \mathcal{L}_l^r , \mathcal{V}_v^r , and \mathcal{T}_t^r , with $r \in \{1, \dots, n_r\}$ are defined. Note that $\mathcal{L}_l^r \cap \mathcal{V}_v^r = \emptyset$, $\mathcal{L}_l^r \cap \mathcal{T}_t^r = \emptyset$, $\mathcal{V}_v^r \cap \mathcal{T}_t^r = \emptyset$, and $\mathcal{L}_l^r \cup \mathcal{V}_v^r \cup \mathcal{T}_t^r = \mathcal{D}_n$ for all $r \in \{1, \dots, n_r\}$.

Then, to select the optimal configuration of hyperparameters \mathcal{H} of the algorithm $\mathcal{A}_{\mathcal{H}}$ in a set of possible ones $\mathfrak{H} = \{\mathcal{H}_1, \mathcal{H}_2, \dots\}$, namely to perform the MS phase, the following procedure has to be applied:

$$\mathcal{H}^* : \arg \min_{\mathcal{H} \in \mathfrak{H}} \sum_{r=1}^{n_r} M(\mathcal{A}_{\mathcal{H}}(\mathcal{L}_l^r), \mathcal{V}_v^r), \quad (4)$$

where $\mathcal{A}_{\mathcal{H}}(\mathcal{L}_l^r)$ is a model learned by \mathcal{A} with the hyperparameters \mathcal{H} based on the the data in \mathcal{L}_l^r and where $M(f, \mathcal{V}_v^r)$ is a desired metric. Since the data in \mathcal{L}_l^r are independent from the ones in \mathcal{V}_v^r , the intuition is that \mathcal{H}^* should be the configuration of hyperparameters which allows achieving optimal performance, according to the desired metric, on a set of data that is independent, namely previously unseen, with respect to the training set.

Then, in order to evaluate the performance of the optimal model, namely the model learned with the optimal hyperparameters based on the available data, which is $f_{\mathcal{A}}^* = \mathcal{A}_{\mathcal{H}^*}(\mathcal{D}_n)$ or, in other words, to perform the EE phase, the following procedure has to be applied:

$$M(f_{\mathcal{A}}^*) = \frac{1}{n_r} \sum_{r=1}^{n_r} M(\mathcal{A}_{\mathcal{H}^*}(\mathcal{L}_l^r \cup \mathcal{V}_v^r), \mathcal{T}_t^r). \quad (5)$$

Since the data in $\mathcal{L}_l^r \cup \mathcal{V}_v^r$ are independent from the ones in \mathcal{T}_t^r , $M(\mathcal{A}_{\mathcal{H}^*}(\mathcal{L}_l^r \cup \mathcal{V}_v^r))$ will be an unbiased estimator of the true performance of the final model [54].

In this paper, the complete k -fold cross-validation is exploited [54, 63] since, together with bootstrap, it represents a state-of-the-art approach to the problem of MD and EE. Then, we need to set $n_r \leq \binom{n}{k} \binom{n-k}{k}$, $l = (k-2) \frac{n}{k}$, $v = \frac{n}{k}$, and $t = \frac{n}{k}$ and the resampling must be done without replacement [63].

Feature Ranking

Once a model is built, namely we perform the learning, MS, and EE phases, it is often required to understand how these models exploit, combine, and extract information in order to understand if the learning process has also cognitive meaning, namely it is able to capture the underlying phenomena and does not just capture spurious correlations [64, 65] by comparing the knowledge of the experts with the information learned by the models. One way to reach this goal is to perform the feature ranking (FR) phase which allows detecting if the importance of those features that are known to be relevant from a physical perspective is also appropriately taken into account, namely ranked as highly important, by the learned models. The failure of the learned model to properly account for the features, which are relevant from a cognitive point of view, might indicate poor quality in the measurements, poor learning ability of the model, or spurious correlations. FR therefore represents a fundamental phase of model checking and verification, since it should generate results consistent with the available knowledge of the phenomena under exam provided by the experts.

FR methods based on RF are one of the most effective FR techniques as shown in many researches [66, 67]. Several measures and approaches are available for FR in RF. One method is based on the permutation test combined with the mean decrease in accuracy (MDA) metric, where the importance of each feature is estimated by removing the association between the feature and outcome of the model. For this purpose, the values of the features are randomly permuted [68] and the resulting increase in error is measured. This way also the influence of the correlated features is also removed. Note that, in our case, as feature we do not intend a particular engineered feature but a particular raw feature (left head inclination, left wrist roundness, etc.), namely all the the features coming from a particular raw feature. More rigorously, for every DT, two main quantities are evaluated: the error on the out-of-bag samples as they are used during prediction and the error on the out-of-bag samples after randomly permuting the values of the features coming from a particular raw feature. The difference between these two values are then averaged

over the different trees in the ensemble and this quantity represents the raw importance score for the variable under exam.

Scenarios

In our experiments, two modelization scenarios have been studied in order to understand the extrapolation capability of the data-driven models:

- Leave One Person Out (LOPO): In this scenario, the model has been trained with all the subjects except one that will be exploited to test the resulting model;
- Leave One Exercise Out (LOEO): In this scenario, the model has been trained with all the the exercises except one that will be exploited to test the resulting model;

Basically the two scenarios just differ in the definitions of \mathcal{L}_l , \mathcal{V}_v , and \mathcal{T}_t , that are the subset of data exploited for building, tuning, and testing the models.

In the LOEO scenario, \mathcal{L}_l , \mathcal{V}_v , and \mathcal{T}_t have been created by randomly selecting data from one exercise to be inserted in \mathcal{T}_t , from another exercise to be inserted in \mathcal{V}_v , and from the remaining ones to be inserted into \mathcal{L}_l .

For the LOPO scenario, we have the same procedure of the LOEO scenario but where the people are considered instead of the exercises.

Experimental Results

This section reports the results of exploiting the methodology presented in “[Methods](#)” on the problem described in “[Data Description](#)” using the data described in the very same section. In all the experiments, we set: $n_r = 100$, $n_t = 1000$, $n_d = \infty$, $n_b = n$, and $n_v \in \{\sqrt[1/3]{n_f}, \sqrt{n_f}, \sqrt[3/4]{n_f}\}$. Experts (violinists 1 to 4) are labeled with $Y = 0$ and beginners (violinists 5 to 7) with $Y = 1$.

Recognition Performances for LOEO and LOPO

Let us present first the recognition performance for the LOPO scenario. Table 3 reports the accuracy on each of the violinists, the overall accuracy, precision, recall, ROC, and ROC-AUC, and the overall confusion matrix. Table 3a shows that some violinists are easier to classify as beginner or experts with respect to the others. Nonetheless, on average, the recognition results are quite high (>70%). More in detail, violinists 1, 2, and 5 are mostly correctly classified having a mean accuracy score very close to 100%. Other violinists, such as violinists 3 and 6, are seldom wrongly classified. Violinists 4 and 7, instead, are mostly wrongly classified. In “[A Comment on the Recognition Performances](#)”, we will give more insights on the reason of

this behavior. The confusion matrix of Table 3b is reported for completeness.

Let us present now the recognition performance for the LOEO scenario. Table 4a, b, c, and d are the counterparts, for the LOEO scenario, of Table 3a, b, c, and d for the LOPO scenario. As one can expect, recognition accuracies in the LOEO scenario are higher than the LOPO ones. In fact, the training phase of the LOEO scenario has more information available with respect to the LOPO one, namely we have to predict if an exercise was performed by an expert or beginner violinist but having the same violinists play other exercises. In particular, it is possible to notice how the scale exercises of violinist 3 are very hard to correctly predict while, for the other two exercises, we obtain a very high accuracy. The expert violinist 4 is easy to be classified when analyzing the scale exercise but it is very hard to distinguish him from beginners in the other two exercises. The accuracy of the scale exercise of violinist 6, instead, has a wide variance. Violinist 7 accuracy predictions are improved with respect to the LOPO scenario but low accuracy scores are achieved in the technique exercise.

A Comment on the Recognition Performances

As we have just observed in “[Recognition Performances for LOEO and LOPO](#)”, some violinists are much more difficult to be correctly classified than others. Experts violinists 1 and 2 are always well classified as expert musicians. The same happens for violinist 5 who is mostly correctly classified as a beginner musician. Instead, we have a very low accuracy for violinist 4 and for violinist 7 in both LOPO and LOEO scenarios. In order to understand these results, we performed a further analysis. The latter consists in observing the original video data with an expert checking what and how the violinist is doing when the algorithm correctly or wrongly classifies the specific violinist.

From the videos, it is possible to observe that the two experts violinists 1 and 2 are often in the same position, with small movements of the pelvis, and with controlled breathing. This behavior led the classifier to perceive that expert violin players have this particular style in the execution of the tracks. The beginner violinist 5, instead, has some hesitations and incertitude during the performance.

From the videos, it is possible to see also that expert violinists 3 and 4 have a different style with respect to the other experts. In fact, they provide much interpretation in their performance, accentuated body sway, and strong breaths exploited to emphasize the beginning of the musical verse. Moreover, violinist 4 has a very peculiar style: he goes up and down on his toes with noticeable vibrations of the head. Violinist 7, on the other side, is a really novel violin player when compared with the other 2 beginners. Difficulties on its prediction are due to the fact that he is

Table 3 (a) Accuracy

Exp/Beg	Violinist	mean	std	min	max
0	1	100.00	0.00	100.00	100.00
	2	99.22	0.86	97.10	100.00
	3	81.72	4.40	72.00	90.00
	4	48.33	6.37	31.48	64.81
1	5	98.05	2.40	89.19	100.00
	6	75.75	9.11	54.16	95.83
	7	10.36	5.44	3.39	32.20
	<i>Mean</i>	<i>73.34</i>	<i>3.22</i>	<i>63.31</i>	<i>83.27</i>

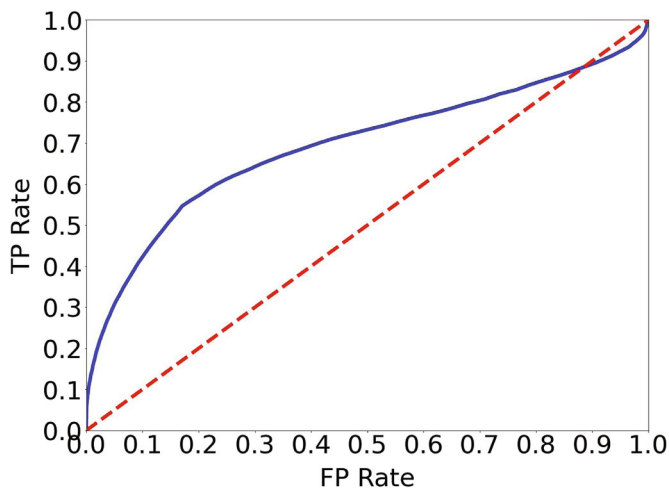
(b) Average confusion matrix (in %) between all exercises and violinists

Prediction	Actual	
	TP <i>50.12</i>	FP <i>10.32</i>
	FN <i>17.92</i>	TN <i>21.64</i>

(c) Average accuracy (%), precision (%), recall (%), and ROC-AUC between all exercises and violinists

Measure	mean	std	min	max
Accuracy	73.34	3.22	63.31	83.27
Precision	67.72	2.80	61.02	74.31
Recall	54.69	3.70	45.83	63.89
ROC-AUC	0.69	0.03	0.62	0.78

(d) Average ROC between all exercises and violinists



concerned about playing the song correctly. From the video, in fact, it is possible to observe how much he is focused on the tracks' execution, with no presence of body sway, with controlled breathing, and with small neck and pelvis movements.

Therefore, the final classifier captures fragility, uncertainties, and hesitations in movements, associating these phenomena with a class of less experienced musicians. Although this is reasonable, the algorithm presents difficulties in understanding what are the essential properties in the execution of a repertoire piece such as, for example, breathing and the emphasis on the strong beats of the music

piece—characteristics that distinguish a more experienced musician. Despite this, an analysis without including sound information can also induce an expert user to not understand the difference between these actions. These differences very well explain the mistakes of the algorithm and the videos have been made freely available⁷ for the convenience of the reader.

Further considerations that can be made observing Table 4a are that there is not an exercise type more

⁷https://www.dropbox.com/sh/uwdvihjfy0uw13/AAD_IXsn-HsH5J0PZ9dGgtK4a

Table 4 (a) LOEO scenario

Exp/Beg	Violinist	Exercise	mean	std	min	max	
0	1	Technique	100.00	0.00	100.00	100.00	
		Repertoire	100.00	0.00	100.00	100.00	
	2	Technique	100.00	0.00	100.00	100.00	
		Scale	91.00	19.31	50.00	100.00	
	3	Repertoire	100.00	0.00	100.00	100.00	
		Technique	100.00	0.00	100.00	100.00	
	4	Scale	50.41	12.99	11.76	82.35	
		Repertoire	99.93	0.48	96.55	100.00	
	1	5	Scale	97.00	5.41	80.00	100.00
			Technique	52.71	15.25	17.65	82.35
		6	Repertoire	33.63	6.67	14.81	51.85
			Scale	98.57	4.30	85.71	100.00
7		Repertoire	100.00	0.00	100.00	100.00	
		Scale	85.00	23.02	50.00	100.00	
Mean		Repertoire	81.57	7.67	56.67	93.33	
		Scale	60.60	14.89	40.00	100.00	
		Technique	26.61	8.29	0.00	50.00	
<i>Mean</i>			<i>80.16</i>	<i>7.53</i>	<i>64.34</i>	<i>91.60</i>	

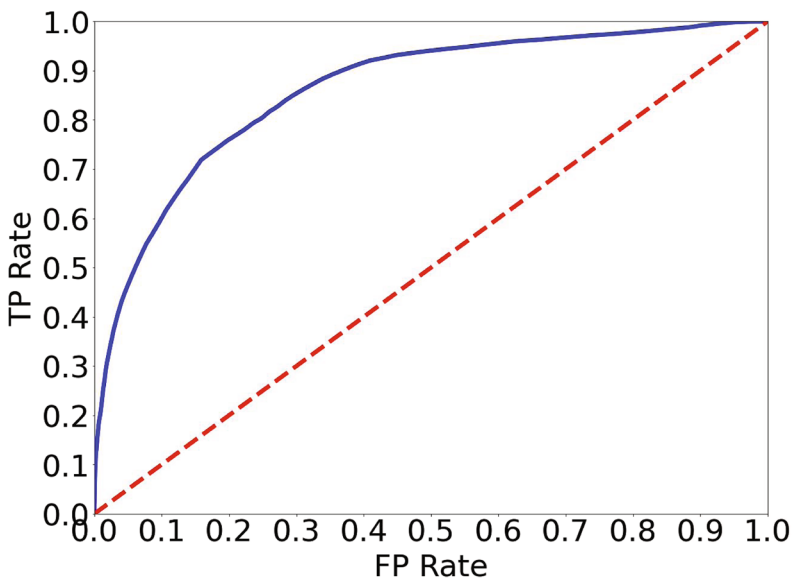
(b) Average confusion matrix (in %) between all exercises and violinists

Prediction	Actual	
	TP 50.85	FP 9.58
FN 11.13	TN 28.44	

(c) Average accuracy (%), precision (%), recall (%), and ROC-AUC between all exercises and violinists

Measure	mean	std	min	max
Accuracy	80.16	7.53	66.34	91.60
Precision	74.80	2.46	68.09	79.59
Recall	71.87	4.07	64.58	81.25
ROC-AUC	0.86	0.02	0.82	0.90

(d) Average ROC between all exercises and violinists



difficult than others to predict correctly or, in other words, to better distinguish experts or beginners. In fact, the accuracies achieved considering only one exercise—technique, repertoire, or scales—are significant enough to allow us to conclude that none of these is better to lead to a more precise classification between experts and beginners. This may seem counter-intuitive because the analysis of the most technical tracks should be alone able to allow this distinction. However, since the analysis only includes movement features, this observation is no longer so restrictive and all types of exercises have the same weight in the final prediction.

Feature Ranking

In both LOPO and LOEO scenarios, we trained our model with features extracted and engineered from the 14 raw data sources discussed in “Data Description”. In order to understand which feature is more relevant in the skill classification of violin players, we applied the method discussed in “Feature Ranking”. Table 5 reports the results of this process and it is easy to notice that the upper body kinetic energy is the most informative feature for our model. This result makes sense since this feature contains information on the whole kinetic part of the upper body, incorporating a lot of information inside it (see “Data Description” for detail). The remaining features are more and more related to parts of the body far away with respect to the influence on the violin, as expected from the comments we reported in the introduction. Observing Table 5, another interesting consideration can be drawn. The left part of the body provides less information with respect to the right one. For example, the right wrist

roundness is more informative than the left wrist roundness. Furthermore, this happens for all features except the left shoulder height. This fact is quite reasonable since the right part of the body is the one involved with the movement of the bow, namely the most dynamic part, while the left side is the one more responsible for the quality of the sound but it has less dynamic. In fact, the bow–violin incidence, which provides further information on the angle that is formed between the two parts of the instrument and has low dynamic but large effect on the quality of the sound, is not considered very important. Furthermore, bow–violin incidence is much more music-dependent than skills-dependent, as it is directly affected by the notes that have to be played. Since in this work we try to understand the violin players’ skill level based on their movements, completely ignoring the sound, it is quite reasonable that parts of the body with higher dynamic are more easy to exploit for this scope with respect to the ones with reduced dynamic. These results support our proposal indicating that the model is not just leaning spurious correlation but is actually understanding the process under exam. A further consideration is that the feature ranking highlights features more sensible to fragilities, uncertainties, and hesitations. This corroborates the considerations discussed in Section 1 on the appropriateness of the algorithm.

Conclusions

In this work, we present a computational approach applied to music education. In particular, a method for classifying the skill level of 7 violinists starting, from data collected from their performances in three exercises properly chosen, in order to better distinguish skill and familiarity in playing the musical instrument. We based our approach on the state-of-art of music education literature and cognitive theories on self-learning, metacognition, and motor learning theory. This approach lets us address the use of a computational approach in a real context. We exploited a state-of-the-art machine learning pipeline including a feature engineering phase guided by the experts of the subject combined with random forests, a state of the art algorithm for classification. Exploiting the intrinsic hierarchy in the dataset, we considered two different extrapolation scenarios, namely on exercises and on violinists, to understand the potentiality of the method. Results show the potentiality of the method but also the necessity of increasing the cardinality of the dataset in terms of both exercises and violinists. In fact, even if the method works as expected and perceives the peculiar characteristics of each exercise and violinist, it sometimes fails and gets deceived by the violinists’ interpretation or focus. We observed how this difficulty is due to a different way to play the musical instrument by each player or simply

Table 5 Feature ranking of the original raw features (from top to least importance)

Rank	Raw feature
1	Upper body kinetic energy
2	Left shoulder height
3	dist lower bow violin
4	Mean shoulder’s velocity
5	Right head inclination
6	Shoulder low back asymmetry
7	Right wrist roundness
8	Right shoulder height
9	dist upper bow violin
10	dist middle bow violin
11	Bow-violin incidence
12	Hand-violin incidence
13	Left wrist roundness
14	Left head inclination

Table 6 Average accuracy (%), precision (%), recall (%), and ROC-AUC between all exercises and violinists for the LOPO and LOEO scenarios and for different models

Scenario	Method	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC
LOPO	RF	73.34 ± 3.22	67.72 ± 2.80	54.69 ± 3.70	0.69 ± 0.03
	Naive Bayes	68.34 ± 4.34	59.98 ± 4.31	51.98 ± 4.92	0.61 ± 0.05
	SVM	70.14 ± 3.98	63.22 ± 3.42	52.12 ± 4.13	0.64 ± 0.04
	Lasso	69.38 ± 4.74	60.19 ± 4.31	50.99 ± 4.35	0.62 ± 0.05
	GSVM	71.54 ± 3.56	65.98 ± 3.38	53.03 ± 3.91	0.67 ± 0.03
	NN	70.65 ± 3.67	64.01 ± 3.85	52.61 ± 4.12	0.65 ± 0.04
LOEO	RF	80.16 ± 7.53	74.80 ± 2.46	71.87 ± 4.07	0.86 ± 0.02
	Naive Bayes	74.04 ± 9.12	66.87 ± 4.15	62.15 ± 5.61	0.78 ± 0.05
	SVM	76.01 ± 8.78	67.82 ± 4.01	64.11 ± 5.32	0.80 ± 0.04
	Lasso	75.64 ± 8.92	67.12 ± 3.95	63.03 ± 5.42	0.79 ± 0.04
	GSVM	78.96 ± 8.93	69.98 ± 3.17	66.75 ± 5.01	0.82 ± 0.04
	NN	77.01 ± 9.03	68.12 ± 3.87	65.10 ± 5.63	0.81 ± 0.03

by atypical movements in the execution. Moreover, a bigger dataset can allow us to better capture these peculiarities in movements of each person, improving the final prediction and providing knowledge about the motor signature of each individual. Increasing the cardinality would also allow the use of more sophisticated tools like the deep learning models able to extract automatically the best set of features needed to lead this classification task.

Nevertheless, this is a first step forward in understand which motion features can be exploited to efficiently distinguish a professional performance and to use such information for real-time student assistive technologies.

Funding Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement. This work is partially supported by the University of Genova through the Bando per l'incentivazione alla progettazione europea 2019 - Mission 1 "Promoting Competitiveness," the EU-H2020-ICT Project TELMI (G.A. 688269), and the EU-H2020-FETPROACT Project ENTIMEMENT (G.A. 824160).

Compliance with Ethical Standards

Conflict of interest Although the authors do not feel any conflict of interest, they declare that this work is partially supported by the University of Genova, "Bando per l'incentivazione alla progettazione europea 2019 - Mission 1 - Promoting Competitiveness," the EU-H2020-ICT Project TELMI (G.A. 688269), and the EU-H2020-FETPROACT Project ENTIMEMENT (G.A. 824160).

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study, formally explicit the use of motion capture, video, and audio recordings for academic research purposes.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Comparison of the Accuracy of Different Algorithms

For completeness, in Table 6, we reported, for both LOPO and LOEO scenarios, the accuracy, the precision, the recall, and the ROC-AUC of different models to show that RF is the best performing algorithm for this problem. In particular, we tested (using the same MS and EE procedures described previously for RF): naive Bayes, linear support vector machines (SVM), SVM with Gaussian kernel (GSVM), lasso, and a multilayer perceptron network trained with stochastic gradient descent (NN). For this purpose, we exploited the R^8 package *caret*⁹ which contains all the previous models: Table 6 shows how RF overperforms all other methods.

⁸<https://www.r-project.org/>

⁹<http://topepo.github.io/caret/index.html>

References

- Hallam S. The development of metacognition in musicians: Implications for education. *Br J Music Educ.* 2001;18(1):27–39.
- McPherson GE, Renwick JM. A longitudinal study of self-regulation in children's musical practice. *Music Educ Res.* 2001;3(2):169–86.
- Pitts S, Davidson J, McPherson G. Developing effective practice strategies. In: *Aspects of teaching secondary music: perspectives on practice*; 2003.
- Leon-Guerrero A. Self-regulation strategies used by student musicians during music practice. *Music Educ Res.* 2008;10(1):91–106.
- Welch GF. Variability of practice and knowledge of results as factors in learning to sing in tune. *Bull Counc Res Music Educ.* 1985;1:238–47.
- Davidson JW. Visual perception of performance manner in the movements of solo musicians. *Psychology of music.* 1993;21(2):103–13.
- Aróstegui JL. *Educating music teachers for the 21st century.* Springer Science & Business Media. 2011.
- Dewey J. *How we think: a restatement of the relation of reflective thinking to the educative process (vol. 8).* Illinois: Carbondale, Illinois: Southern Illinois University Press; 1933.
- Kluwe RH. Executive decisions and regulation of problem solving behavior. *Metacognition, motivation, and understanding.* 1987;2:31–64.
- Hacker DJ, Dunlosky J, Graesser AC. *Metacognition in educational theory and practice.* Routledge. 1998.
- Flavell JH. *Metacognitive aspects of problem solving. The nature of intelligence.* 1976.
- Fry PS, Lupart JL. *Cognitive processes in children's learning: practical applications in educational practice and classroom management.* Charles C Thomas Pub Limited. 1987.
- Barry NH. The effects of practice strategies, individual differences in cognitive style, and gender upon technical accuracy and musicality of student instrumental performance. *Psychol Music.* 1992;20(2):112–23.
- Nielsen S. Self-regulating learning strategies in instrumental music practice. *Music education research.* 2001;3(2):155–67.
- Brandfonbrener AG. Musculoskeletal problems of instrumental musicians. *Hand clinics.* 2003;19(2):231–9.
- Fishbein M, Middlestadt SE, Ottati V, Straus S, Ellis A. Medical problems among icom musicians: overview of a national survey. *Medical problems of performing artists.* 1988;3(1):1–8.
- Magill R, Anderson D. *Motor learning and control.* New York: McGraw-Hill Publishing; 2010.
- Ballreich R, Baumann W. *Grundlagen der biomechanik des sports.* In: *Probleme, Methoden, Modelle.* Stuttgart: Enke; 1996.
- Marquez-Borbon A. Perceptual learning and the emergence of performer-instrument interactions with digital music systems. In: *Proceedings of a body of knowledge - embodied cognition and the arts conference*; 2018.
- Visentin P, Shan G, Wasiak EB. Informing music teaching and learning using movement analysis technology. *Int J Music Educ.* 2008;26(1):73–87.
- Volta E, Mancini M, Varni G, Volpe G. Automatically measuring biomechanical skills of violin performance: an exploratory study. In: *International conference on movement and computing*; 2018.
- Askenfelt A. Measurement of the bowing parameters in violin playing. ii: Bow-bridge distance, dynamic range, and limits of bow force. *The Journal of the Acoustical Society of America.* 1989;86(2):503–16.
- Shalev-Shwartz S, Ben-David S. *Understanding machine learning: From theory to algorithms.* Cambridge: Cambridge University Press; 2014.
- Goodfellow I, Bengio Y, Courville A. *Deep learning.* Cambridge: MIT press; 2016.
- Dalmazzo DC, Ramírez R. Bowing gestures classification in violin performance: a machine learning approach. *Frontiers in psychology.* 2019;10:344.
- Li W, Pasquier P. Automatic affect classification of human motion capture sequences in the valence-arousal model. In: *International symposium on movement and computing*; 2016.
- Kapsouras I, Nikolaidis N. Action recognition on motion capture data using a dynemes and forward differences representation. *J Vis Commun Image Represent.* 2014;25(6):1432–45.
- Wang JM, Fleet DJ, Hertzmann A. Gaussian process dynamical models for human motion. *IEEE transactions on pattern analysis and machine intelligence.* 2007;30(2):283–98.
- Peiper C, Warden D, Garnett G. An interface for real-time classification of articulations produced by violin bowing. In: *Conference on new interfaces for musical expression*; 2003.
- Cho K, Chen X. Classifying and visualizing motion capture sequences using deep neural networks. In: *International conference on computer vision theory and applications*; 2014.
- Butepage J, Black MJ, Kragic D, Kjellstrom H. Deep representation learning for human motion prediction and classification. In: *IEEE conference on computer vision and pattern recognition*; 2017.
- Al-Radaideh QA, Bataineh DQ. A hybrid approach for arabic text summarization using domain knowledge and genetic algorithms. *Cognitive Computation.* 2018;10(4):651–69.
- Oliva J, Serrano JI, DelCastillo MD, Iglesias A. Cross-linguistic cognitive modeling of verbal morphology acquisition. *Cognitive Computation.* 2017;9(2):237–58.
- Keuninckx L, Danckaert J, Vander Sande G. Real-time audio processing with a cascade of discrete-time delay line-based reservoir computers. *Cognitive Computation.* 2017;9(3):315–26.
- Wang H, Xu L, Wang X, Luo B. Learning optimal seeds for ranking saliency. *Cognitive Computation.* 2018;10(2):347–58.
- Zhang HG, Wu L, Song Y, Su CW, Wang Q, Su F. An online sequential learning non-parametric value-at-risk model for high-dimensional time series. *Cognitive Computation.* 2018;10(2):187–200.
- Wang B, Zhu R, Luo S, Yang X, Wang G. H-mrst: a novel framework for supporting probability degree range query using extreme learning machine. *Cognitive Computation.* 2017;9(1):68–80.
- Scardapane S, Uncini A. Semi-supervised echo state networks for audio classification. *Cognitive Computation.* 2017;9(1):125–35.
- Huang GB. An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation.* 2014;6(3):376–90.
- Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
- Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems?. *The Journal of Machine Learning Research.* 2014;15(1):3133–81.
- Wainberg M, Alipanahi B, Frey BJ. Are random forests truly the best classifiers?. *The Journal of Machine Learning Research.* 2016;17(1):3837–41.
- Galton F. *Vox populi.* Nature Publishing Group. 1907.
- Liu N, Sakamoto JT, Cao J, Koh ZX, Ho AFW, Lin Z, Ong MEH. Ensemble-based risk scoring with extreme learning machine for prediction of adverse cardiac events. *Cognitive Computation.* 2017;9(4):545–54.
- Wen G, Hou Z, Li H, Li D, Jiang L, Xun E. Ensemble of deep neural networks with probability-based fusion for facial expression recognition. *Cognitive Computation.* 2017;9(5):597–610.

46. Bosse T, Duell R, Memon ZA, Treur J, Van DerWal CN. Agent-based modeling of emotion contagion in groups. *Cognitive Computation*. 2015;7(1):111–36.
47. Ortín S, Pesquera L. Reservoir computing with an ensemble of time-delay reservoirs. *Cognitive Computation*. 2017;9(3):327–36.
48. Cao L, Sun F, Liu X, Huang W, Kotagiri R, Li H. End-to-end convnet for tactile recognition using residual orthogonal tiling and pyramid convolution ensemble. *Cognitive Computation*. 2018;10(5):718–36.
49. Li Y, Zhu E, Zhu X, Yin J, Zhao J. Counting pedestrian with mixed features and extreme learning machine. *Cognitive Computation*. 2014;6(3):462–76.
50. Ofek N, Poria S, Rokach L, Cambria E, Hussain A, Shabtai A. Unsupervised commonsense knowledge enrichment for domain-specific sentiment analysis. *Cognitive Computation*. 2016;8(3):467–77.
51. Volpe G, Kolykhalova K, Volta E, Ghisio S, Waddell G, Alborno P, Piana S, Canepa C, Ramirez-Melendez R. A multimodal corpus for technology-enhanced learning of violin playing. In: Biannual conference on Italian SIGCHI chapter; 2017.
52. Camurri A, Coletta P, Varni G, Ghisio S. Developing multimodal interactive systems with eyesweb xmi. In: International conference on new interfaces for musical expression; 2007.
53. Aggarwal CC. *Data mining: the textbook*. Berlin: Springer; 2015.
54. Oneto L. *Model selection and error estimation in a nutshell*. Berlin: Springer; 2019.
55. VanLaerhoven K, Cakmakci O. What shall we teach our pants?. In: International symposium on wearable computers; 2000.
56. DeVaul RW, Dunn S. Real-time motion classification for wearable computing applications. MIT Technical Report. 2001.
57. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: International workshop on ambient assisted living; 2012.
58. Sama A, Pardo-Ayala DE, Cabestany J, Rodríguez-Molinero A. Time series analysis of inertial-body signals for the extraction of dynamic properties from human gait. In: International joint conference on neural networks; 2010.
59. Wang N, Ambikairajah E, Lovell NH, Celler BG. Accelerometry based classification of walking patterns using time-frequency analysis. In: IEEE engineering in medicine and biology society; 2007.
60. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: International conference on pervasive computing; 2004.
61. Rokach L, Maimon OZ, Vol. 69. *Data mining with decision trees: Theory and applications*. Singapore: World Scientific; 2008.
62. Orlandi I, Oneto L, Anguita D. Random forests model selection. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN); 2016.
63. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International joint conference on artificial intelligence; 1995.
64. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3:1157–82.
65. Calude CS, Longo G. The deluge of spurious correlations in big data. *Foundations of science*. 2017;22(3):595–612.
66. Saeys Y, Abeel T, Vande Peer Y. Robust feature selection using ensemble feature selection techniques. In: Joint European Conference on machine learning and knowledge discovery in databases; 2008.
67. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. *Pattern Recogn Lett*. 2010;31(14):2225–36.
68. Good P. *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Berlin: Springer Science & Business Media; 2013.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.