

Intégration de l'IA et de la biomécanique pour la prévention des blessures sportives

Estelle Delouche

Talan

Les blessures chez les sportifs de haut niveau représentent un enjeu majeur tant sur le plan physiologique que socio-économique. Elles impliquent des périodes de convalescence prolongées, une diminution du niveau de performance individuelle et collective, ainsi qu'un impact financier important pour les clubs professionnels.

La prévention des blessures pour les athlètes est donc devenue un axe de recherche important pour les clubs, mobilisant des équipes scientifiques pluridisciplinaires qui étudient à la fois des données physiologiques classiques (fréquence cardiaque «*Fc*» ou la $VO_2\text{max}$) mais également les cycles d'activation musculaire en analysant les électromyogrammes (EMG).

Le projet de recherche présenté dans cet article résulte d'une collaboration entre l'ENSAM, le club de rugby du Stade français Paris, le centre de recherche et innovation Talan, et s'inscrit dans une démarche de prévention active des blessures, plus spécifiquement la rupture du ligament croisé antérieur (LCA), une lésion fréquente au niveau du genou.

Nous cherchons à déterminer si des schémas d'activation musculaire associés à un risque accru de rupture du ligament croisé antérieur sont identifiables dans les électromyogrammes.

Pourquoi étudier les EMG ?

Les signaux EMG enregistrent les phases de contraction et de relaxation musculaire, et offrent un aperçu de la coordination neuromusculaire. Ainsi nous pourrions déchiffrer les informations de ces signaux pour déterminer si une forme d'onde est représentative d'un mouvement à risque (un changement de direction dangereux pour le tendon par exemple). Cependant, comprendre les signaux EMG est une tâche complexe : ils sont également composés de bruit et d'artefacts qui couvrent les informations et compliquent les interprétations.

Dans ce contexte, l'électromyographie de surface (EMG) apparaît comme un outil de choix pour étudier les stratégies d'activation neuromusculaire susceptibles de prévenir ou, à l'inverse, de favoriser la survenue de lésions du LCA [8, 13, 7]. Introduite par Lemman et Ritchie en 1979, l'EMG de surface est une technique non invasive qui permet d'enregistrer, à l'aide d'électrodes placées directement sur la peau, l'activité électrique produite par les muscles lors de leur contraction.

L'analyse de ces signaux aide à la compréhension du système neuromusculaire en évaluant l'intensité et la coordination des activations des différents muscles. Dans le cadre de notre étude, un enregistrement simultané de plusieurs groupes musculaires a été réalisé, afin de cartographier la dynamique d'activation autour de l'articulation du genou. L'observation conjointe des signaux EMG des quadriceps et des ischio-jambiers permet ainsi d'identifier les contributeurs principaux au mouvement, et d'isoler les motifs d'activation potentiellement à risque, en particulier lors des phases critiques précédant un changement de direction [5].

Ainsi, le projet vise à développer une méthodologie d'analyse de ces signaux, dans le but d'identifier d'éventuels motifs d'activation musculaire associés à un risque accru de blessure.

Pourquoi utiliser l'intelligence artificielle ?

Les signaux EMG sont très riches en information sur l'activité musculaire, mais présentent également une forte complexité temporelle, une variabilité interindividuelle élevée et une présence importante de bruit. Les méthodes de statistiques classiques atteignent alors rapidement leurs limites. C'est dans ce contexte que l'intelligence artificielle, et plus particulièrement les techniques de *machine learning* et *deep learning*, apparaissent comme des outils prometteurs [15, 11, 9, 4].

Au-delà de la simple extraction de caractéristiques statistiques, l'intelligence artificielle offre un cadre pour modéliser les signaux EMG en tant que séries temporelles complexes. Les réseaux de neurones convolutifs (CNN)

permettent de détecter automatiquement des motifs locaux dans le temps, en apprenant des filtres capables de représenter les phases de contraction ou d'impact musculaire. Les architectures récurrentes, telles que les LSTM (*long short-term memory*) ou les GRU (*gated recurrent units*), facilitent quant à elles, la prise en compte des dépendances temporelles longues et des interactions dynamiques entre muscles. L'intégration de ces modèles dans une chaîne d'apprentissage supervisé permet de transformer l'analyse biomécanique en une tâche de reconnaissance de motifs temporels, où la performance est évaluée à l'aide d'indicateurs quantitatifs standard (taux de classification, précision, rappel, F1-score, ou encore AUC). Cette approche algorithmique vise à identifier des régularités latentes dans des signaux bruités et non stationnaires, tout en améliorant la robustesse et la possibilité de généraliser des modèles prédictifs.

Ainsi ces techniques sont particulièrement puissantes pour extraire automatiquement des motifs complexes à partir de données multivariées et bruitées et peuvent détecter des signaux faibles dans les signaux EMG [8, 12].

L'objectif de cette recherche est donc d'évaluer, à travers une approche combinant l'analyse de signaux EMG et des techniques d'intelligence artificielle, dans quelle mesure il est possible de discriminer des motifs d'activation musculaire potentiellement à risque, et de contribuer ainsi à une meilleure compréhension et anticipation des mécanismes opérant lors d'une LCA.

Comment surviennent les lésions du ligament croisé antérieur ?

Sur un plan anatomique, le ligament croisé antérieur s'insère sur la partie inférieure du fémur, traverse obliquement l'articulation du genou, puis vient se fixer sur la partie supérieure du tibia. Il forme avec le ligament croisé postérieur (LCP), une structure dite « croisée » qui permet de stabiliser le genou lors des mouvements de flexion, d'extension et de rotation. Principalement, il sert à protéger le genou pour restreindre la rotation ou translation antérieure et éviter la lésion.

Sa rupture correspond à une déchirure partielle ou totale du ligament, souvent à ses extrémités d'insertion, et résulte le plus souvent d'un traumatisme indirect [6, 2]. Dans le cas du rugby, les études ont montré que la majorité des ruptures du LCA surviennent sans contact direct avec d'autres joueurs, mais lors de mouvements brusques tels que des changements de direction rapides. Plus généralement cette rupture survient dans les 50 ms suivant l'impact du pied au sol [16].

Sur un plan biomécanique, deux facteurs ont été identifiés comme déterminants dans les ruptures du LCA :

- un déséquilibre du ratio de coactivation musculaire (notamment entre les quadriceps et les ischio-jambiers);
- un retard d'activation musculaire, ne permettant pas de stabiliser efficacement l'articulation lors de la phase critique d'appui.

Ces mécanismes conduisent à une surcharge articulaire non compensée, pouvant aboutir à une rupture ligamentaire [3], qui ne cicatrise généralement pas spontanément et requiert une intervention chirurgicale dans la majorité des cas.

Collecte de données et scénarios d'acquisition

Ainsi dans le cadre de notre projet nous souhaitons analyser l'activation musculaire des muscles qui protègent le genou afin de déterminer si un ou plusieurs enregistrent des phases dites à risque lors d'un changement de direction soudain. Pour cela des électrodes ont été placées sur deux groupes musculaires : les quadriceps et les ischio-jambiers. Un total de 5 muscles sont monitorés : le rectus femoris, le vastus lateralis, le vastus medialis, le biceps femoris et le semitendinosus (figure 1).



Fig. 1. Position des différents muscles observés dans notre étude. De *Bansal et al. [1].

Afin d'analyser les schémas d'activation musculaire dans des conditions réalistes, deux scénarios expérimentaux ont été conçus par les ingénieurs en biomécanique de l'ENSAM (figure 2). Le premier scénario, désigné dans la suite de l'article comme scénario *planifié*, consiste en une course linéaire de 15 mètres, au cours de laquelle le joueur de rugby récupère le ballon en mouvement avant d'effectuer un changement de direction prédéfini (à gauche

ou à droite), communiqué en amont de l'essai. Ce protocole vise à reproduire un déplacement maîtrisé dans un environnement contrôlé.

À l'inverse, le scénario *non-planifié*, également qualifié de scénario avec *incertitude*, vise à reproduire les conditions imprévisibles du jeu réel. Dans ce cas, le joueur ne connaît pas à l'avance la direction à prendre en fin de course. Pour simuler cette situation d'incertitude, un système de déclenchement visuel en temps réel a été mis en place. Lorsqu'un joueur franchit une paire de repères visuels (piliers jaunes) situés sur le parcours, un signal visuel instantané (flèche gauche ou droite) s'affiche sur un écran placé en face de lui, indiquant la direction à suivre immédiatement. Ce dispositif réduit le temps de prise d'information, contraignant le joueur à une prise de décision rapide, proche des exigences du jeu en situation réelle.

L'ensemble de ce protocole a été appliqué à un échantillon de 11 joueurs, chacun réalisant plusieurs essais dans les deux conditions. Au total, 1980 signaux EMG ont été enregistrés, fournissant une base de données pour l'application des algorithmes de *machine learning* et *deep learning*.

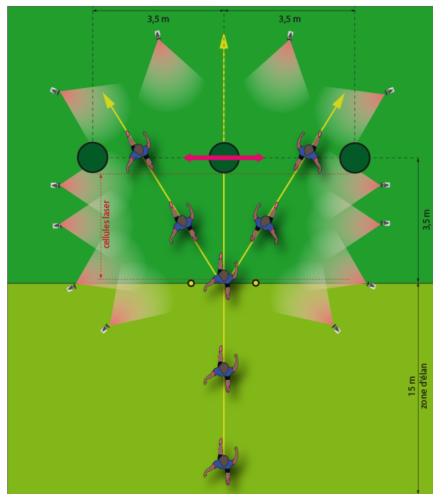


Fig. 2. Schéma du dispositif mis en place par l'ENSAM pour l'acquisition des données.

Pre-processing et processing des EMG

Les signaux électromyographiques (EMG) enregistrés nécessitent une phase de prétraitement et de traitement afin de garantir la fiabilité des analyses ultérieures. Dans un premier temps, le prétraitement vise à nettoyer et structurer les données brutes. Cela inclut l'identification et la suppression

des données aberrantes (figure 3). Le signal est ensuite centré (*demeaned*) pour éliminer toute composante continue résiduelle. Enfin, les portions de signal manquantes sont reconstruites par imputation temporelle locale, en les remplaçant par une fenêtre de même taille issue d'un segment antérieur immédiat du même signal, assurant ainsi la continuité et la cohérence temporelle et fréquentielle.

Concernant le traitement, un filtrage passe-bande entre 20 Hz et 450 Hz est appliqué afin de supprimer les artefacts de basse fréquence et les interférences haute fréquence [14]. Une normalisation du signal est ensuite envisagée, souvent par rapport à une mesure biomécanique de référence (par exemple, le pic de couple musculaire), afin de comparer les niveaux d'activation entre individus.

Dans la suite de cet article, nous présentons les résultats des algorithmes de *machine learning* sur des signaux EMG prétraités et filtrés mais où aucune normalisation n'a été appliquée.

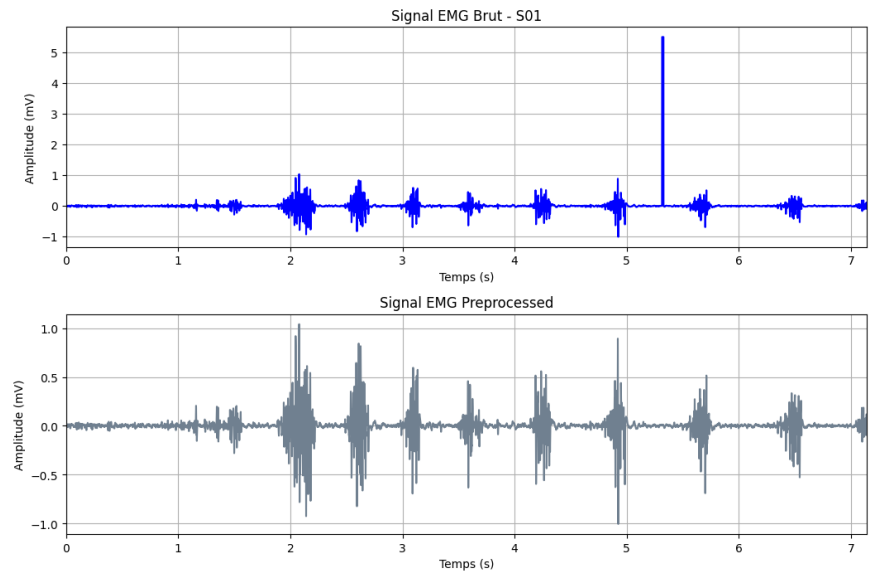


Fig. 3. De haut en bas : signal EMG brut et signal EMG après traitement (la donnée aberrante a été supprimée et le signal filtré).

Analyse des signaux EMG par apprentissage supervisé

Extraction de caractéristiques et détection de variations

Nous avons exploré la capacité des techniques d'apprentissage supervisé à discriminer les signaux EMG issus de scénarios planifiés de ceux avec incertitude afin de déterminer si ces deux conditions induisent des motifs musculaires différenciables.

Pour cela, les signaux EMG ont été segmentés en fenêtres temporelles de 300 échantillons (soit environ 140 ms), avec un recouvrement de 50 %, afin de garantir une continuité dans l'analyse. Sur chacune de ces fenêtres, nous avons extrait un ensemble de caractéristiques tels que des descripteurs statistiques (moyenne, écart-type, kurtosis, *skewness*), des mesures d'énergie (entropies de Shannon et de Rényi), ainsi que des descripteurs de formes caractérisant la morphologie locale du signal.

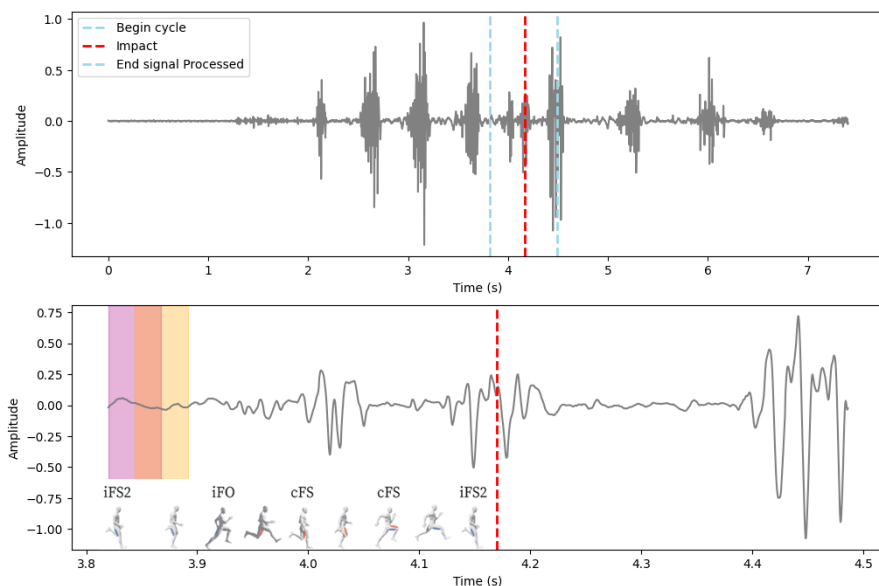


Fig. 4. Panel supérieur : exemple d'un signal EMG avec l'indication de la zone d'impact en lignes pointillées bleues. Panel inférieur : zoom sur la zone d'impact avec indication en bas à gauche du cycle de course (repris de [10]). Les rectangles colorés indiquent les fenêtres d'analyses du signal : 47 ms de longueur avec un overlap de 50%.

Cependant, la dynamique globale étant fortement bruitée et difficile à interpréter, nous avons choisi de restreindre l'analyse à une zone d'intérêt physiologiquement critique (figure 4), centrée autour de l'impact initial du pied au sol précédant le changement de direction. Cette fenêtre temporelle de 0,7 seconde nous permet de centrer nos recherches de signaux autour de la zone la plus à risque. Dans cette configuration restreinte, les signaux ont été segmentés en fenêtres plus fines de 100 échantillons (soit 47 ms), ce qui correspond à l'échelle temporelle typique d'apparition du risque de lésion ligamentaire (environ 50 ms après l'impact final [16]).

L'ensemble des caractéristiques extraites dans cette zone a ensuite été utilisé pour alimenter des algorithmes de classification supervisé (*random forest*, *gradient boosting*, *decision tree*), dans le but de détecter l'existence éventuelle de sous-groupes de signaux différenciant les deux scénarios expérimentaux.

Les résultats indiquent des bonnes classifications de l'ordre de 72 % comme l'indique la matrice de confusion (figure 5).

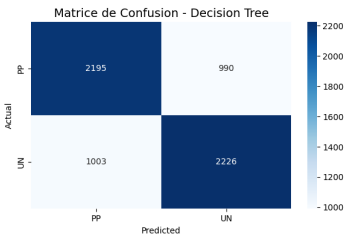


Fig. 5. Matrice de confusion calculée sur les résultats de *decision tree*. «PP» correspond aux signaux provenant d'un scénario planifié et «UN» fait référence au scénario avec incertitude.

Peut-on améliorer les prédictions ?

Compte tenu des performances limitées des algorithmes de classification supervisée, dont les taux de prédiction n'excèdent pas 72 %, nous avons exploré une approche complémentaire fondée sur la labellisation explicite des phases d'activation musculaire. L'objectif est de mieux caractériser la dynamique locale des signaux EMG, en distinguant les différentes phases physiologiques associées au mouvement, et en particulier la phase d'impact. Ainsi, nous avons attribué un label correspondant à 3 états possibles (figure 6) :

1. «repos» (non-contraction);
2. contraction musculaire active;

3. impact, correspondant à la phase d'appui au sol précédant immédiatement un changement de direction.

La méthode de labellisation repose sur le calcul de la RMS (*root mean square*) du signal EMG. Un seuil empirique a été fixé pour discriminer les phases de repos (valeurs RMS inférieures au seuil) des phases de contraction (valeurs RMS supérieures). La zone d'impact, quant à elle, a été définie à partir d'annotations temporelles précises alignées sur les données cinématiques (moment du contact initial du pied au sol), et reçoit une étiquette dédiée. Cette stratégie permet de structurer les données EMG selon leur contenu fonctionnel et à extraire précisément les caractéristiques de chaque phase.

Ce travail de labellisation est actuellement toujours en cours mais nous pensons que les résultats pourront nous donner une meilleure compréhension de la dynamique des muscles et de mettre en évidence des aspects caractéristiques à chaque phase.

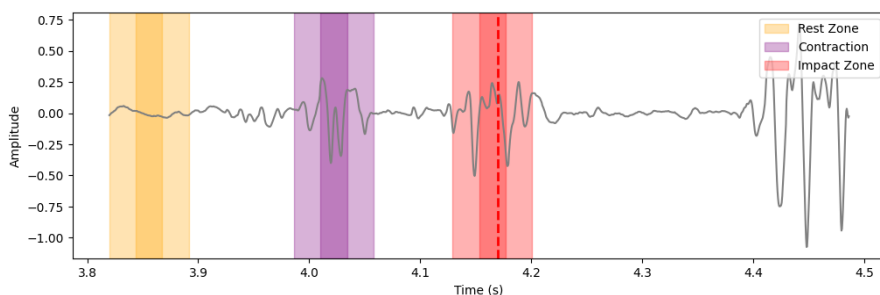


Fig. 6. Labellisation d'un signal selon le niveau de la RMS. En jaune, muscle considéré au repos, en violet en contraction et, finalement, rouge correspond à la zone d'impact.

Défis méthodologiques liés à l'analyse des signaux EMG

L'analyse et la modélisation des signaux électromyographiques se heurtent à plusieurs contraintes structurelles et physiologiques qui complexifient la détection de motifs généralisables :

1. un premier défi réside dans la variabilité temporelle de l'activation musculaire : même pour une tâche motrice identique, le moment d'activation des muscles peut différer d'un essai à l'autre, perturbant l'alignement temporel des signaux et rendant difficile la comparaison directe entre

cycles; cette instabilité est exacerbée par les différences physiologiques interindividuelles, telles que la masse musculaire, le niveau de coordination neuromusculaire ou encore l'état de fatigue, qui modifient l'intensité et la dynamique des signaux enregistrés;

2. le nombre de muscles étudiés : l'analyse des données EMG dans cette étude repose sur l'enregistrement simultané de cinq muscles par essai, soit un total de 1980 signaux; cette granularité, bien qu'indispensable pour capturer la coordination musculaire autour du genou, engendre une complexité supplémentaire dans l'interprétation des résultats; en effet, chaque muscle présente un motif d'activation propre, avec des temporalités et des rôles fonctionnels distincts dans le mouvement; cela signifie que les données ne sont pas indépendantes, mais interconnectées dans une dynamique neuromusculaire globale, ce qui rend l'analyse plus complexe;
3. enfin, la taille restreinte du jeu de données : l'étude repose sur un échantillon de 11 joueurs, tous issus de postes différents sur le terrain (pilier, talonneur, ailier, etc.), ce qui implique des profils physiologiques hétérogènes en lien avec les spécificités de leur rôle; or, ces différences, en termes de morphologie ou de masse musculaire, ne sont pas prises en compte dans les analyses actuelles; cela a pour effet de réduire la performance des algorithmes d'apprentissage et limite la possibilité d'identifier un motif d'activation commun ou caractéristique d'une situation à risque.

Conclusion

Ce projet de recherche, mené conjointement par l'ENSAM, le Stade Français Paris et le centre R&D de Talan, rassemble une équipe pluridisciplinaire de scientifiques afin d'identifier des motifs d'activation musculaire caractéristiques dans les signaux EMG.

Les premiers résultats montrent que l'extraction de caractéristiques permet déjà de distinguer les signaux issus de scénarios planifiés de ceux en « incertitude », avec un taux de bonne classification atteignant 72 %. Cependant, afin de mieux comprendre ces mécanismes, nous effectuons une labellisation des données afin d'extraire plus précisément les aspects caractéristiques des cycles musculaires et de la zone d'impact. À terme, nous souhaitons étendre cette analyse à des approches basées sur le *deep learning*, en mobilisant des architectures de réseaux neuronaux adaptées aux séries temporelles (telles que les CNN et les LSTM), afin de surmonter la forte variabilité intra et inter-individuelle des signatures d'activation musculaire.

Références

- [1] H. Bansal, B. Chinagundi, P.S. Rana & N. Kumar. 2022. An ensemble machine learning technique for the detection of abnormalities in knee movement sustainability. *Sustainability*, 14(20), 13464.
- [2] N.A. Bates, N.D. Schilaty, A.J. Krych & T.E. Hewett. 2019. Variation in ACL and MCL strain before initial contact is dependent on injury risk level during simulated landings. *Orthopaedic journal of sports medicine*, 7(11), 2325967119884906.
- [3] T.F. Besier, D.G. Lloyd & T.R. Ackland. 2003. "Muscle Activation Strategies at the Knee during Running and Cutting Maneuvers;," *Medicine & Science in Sports & Exercise*, vol. 35, no. 1, pp. 119–127, DOI:10.1097/00005768-200301000-00019.
- [4] L. Cervoni, R. Sleiman, D. Jacob & M. Roudesli. 2023. Explainable Artificial Intelligence in Response to the Failures of Musculoskeletal Disorder Rehabilitation. In *International Workshop on Explainable Artificial Intelligence in Healthcare* (pp. 14–24). Cham: Springer Nature Switzerland.
- [5] S. Colby, A. Francisco, Y. Bing, D. Kirkendall, M. Finch & W. Garrett. 2000. Electromyographic and kinematic analysis of cutting maneuvers: implications for anterior cruciate ligament injury. *The American journal of sports medicine*, 28(2), 234–240.
- [6] F. Della Villa, M. Buckthorpe, A. Grassi, A. Nابیuzzi, F. Tosarelli, S. Zaffagnini & S. Della Villa. 2020. Systematic video analysis of ACL injuries in professional male football (soccer): injury mechanisms, situational patterns and biomechanics study on 134 consecutive cases. *British journal of sports medicine*, 54(23), 1423–1432.
- [7] C.J. De Luca. 2002. Surface electromyography: Detection and recording. *DelSys Incorporated*, 10(2), 1–10.
- [8] M.E. Hahn. 2007. "Feasibility of estimating isokinetic knee torque using a neural network model," *Journal of Biomechanics*, vol. 40, no. 5, pp. 1107–1114, DOI:10.1016/j.jbiomech.2006.04.014.
- [9] D. Jacob, R. Tievant, L. Cervoni & M. Roudesli. 2023. Prédiction des blessures au Foot 5 à l'aide d'une méthode de machine learning. *Journal de Traumatologie du Sport*, 40(4), 261–269.
- [10] G. Kakehata, Y. Goto, S. Ido, and K. Kanosue. 2021. Timing of Rectus Femoris and Biceps Femoris Muscle Activities in Both Legs at Maximal Running Speed. *Med. Sci. Sports Exerc.*, Vol. 53, No. 3, pp. 643–652.
- [11] B. Karlik. 2014. Machine learning algorithms for characterization of EMG signals. *International Journal of Information and Electronics Engineering*, 4(3), 189.
- [12] S.H. Park & S.P. Lee. 1998. EMG pattern recognition based on artificial intelligence techniques. *IEEE transactions on Rehabilitation Engineering*, 6(4), 400–405.
- [13] J. Sun, Y. Wang, J. Hou, G. Li, B. Sun & P. Lu. 2024. Deep Learning for Electromyographic Lower-Limb Motion Signal Classification Using Residual Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 2078–2086. <https://doi.org/10.1109/TNSRE.2024.3403723>.
- [14] J. Wang, L. Tang & J.E. Bronlund. 2013. Surface EMG signal amplification and filtering. *International Journal of Computer Applications*, 82(1).
- [15] J. Yousefi & A. Hamilton-Wright. 2014. Characterizing EMG data using machine-learning tools. *Computers in biology and medicine*, 51, 1–13.
- [16] M.K. Zebis, J. Bencke, L.L. Andersen, S. Døssing, T. Alkjær, S. P. Magnusson, M. Kjær & P. Aagaard. 2008. The effects of neuromuscular training on knee joint motor control during sidcutting in female elite soccer and handball players. *Clinical journal of sport medicine*, 18(4), 329–337.

