Conference Paper

Design of A Normative sEMG Database for Biometric Comparison in Rehabilitation Research

Athanasios Arvanitidis[*], Konstantinos Mitsopoulos, Vasiliki Fiska, Alkinoos Athanasiou and Panagiotis D. Bamidis

Aristotle University of Thessaloniki, School of Medicine, Lab of Medical Physics and Digital Innovation, Thessaloniki, Greece.

[*] Corresponding Author Email: thanosarv99@gmail.com

ABSTRACT

Electromyography (EMG) is used in a wide range of research fields, such as physiotherapy, ergonomics, and neurorehabilitation. Normative EMG databases play a crucial and significant role in the efficient diagnosis and treatment of neuromuscular disorders. They can rapidly provide information that, although not necessarily diagnostic, can efficiently and effectively guide further diagnostic studies. Quantitative electromyography (QEMG) in the upper extremities is an effective diagnostic tool, but there are currently few normative databases available. The absence of fundamental guidelines and established methods for creating normative databases contributes to a significant obstacle in the field of rehabilitation research. This study aims to bridge this gap by designing a dynamic, scalable, consistent, available, and partition-tolerant NoSQL database (DB), in alignment with the Consistency, Availability, and Partition Tolerance (CAP) theorem, to house normative surface electromyography (sEMG) values for upper body muscles, primarily for biometric comparison in rehabilitation. The DB encompasses diverse EMG features, both in the time and frequency domains, as well as anthropometric variables, extracted by healthy participants and post-stroke or spinal cord injury patients. The participant selection is based on Greece's average demographic statistics and specific inclusion and exclusion criteria from existing clinical trials. The proposed DB is particularly designed to be continuously updated offering real-time insights, allowing the DB to be an even more valuable resource for researchers and practitioners working in the field.

Keywords—*Quantitative electromyography, Rehabilitation, Biomedical database, NoSQL*.

INTRODUCTION

The electrochemical and mechanical activities that occur during the biological events of the human body frequently generate measurable signals, known as biosignals, that can be analyzed. These signals offer valuable insights into not only the intrinsic physiological and pathophysiological states of the body, but also into an individual's affective, attentional, and other cognitive states. Such information is fundamental in understanding the underlying mechanisms of specific biological systems or events and holds significant potential for medical diagnosis.[1]

EMG is a specific biosignal generated from electrophysiological changes in muscle fiber membrane conductivity and quantifies the electrical currents produced during muscle contractions.[2,3] EMG sensors can generate information relevant to muscular activity and are widely used in clinical settings for diagnosing conditions and diseases of the central and peripheral nervous systems that involve the sensorimotor and somatosensory pathways.

Quantitative Electromyography (QEMG), offers a valuable approach to diagnostics, shedding light on neural and muscular disorders. However, a solemn omission in the current landscape is the lack of a comprehensive normative database (DB) that can serve as a standard for biometric comparison.[4] This deficiency has far-reaching implications, particularly in rehabilitation research, where effective diagnosis and treatment are contingent upon comparative analyses.

Surface Electromyography (sEMG) offers a non-invasive yet robust way to acquire useful information about the human body's physiological and pathophysiological states. Given its non-invasive nature, sEMG has been considered an invaluable tool for studying a myriad of conditions, ranging from genetic neuromuscular disorders to spinal cord injury.[5] In recent years, advancements in sEMG technologies have positioned them as a complement or even a potential alternative to needle Electromyography (nEMG) and Nerve Conduction Studies (NCS). However, this technological leap is undermined by several limitations.[6] This study aims to tackle the lack of standardized DBs for sEMG data, which restricts their comparability and, as a result, hampers their clinical utility.[4,7–9]

The current study aims to fill this void, by creating a NoSQL DB with normative sEMG values of the upper body. The CAP theorem has been more known in recent years as a crucial framework for understanding the limitations and potential trade-offs in developing distributed DBs.[10] The theorem states that only two of the following three properties—consistency, which ensures that all data replicas are synchronized, high availability, which ensures uninterrupted access to data for updates, and partition tolerance, which allows for continued operation in the face of network failures—can be maintained by such a system at an optimal level.[11] Leveraging the capabilities of MongoDB, the DB is designed to be dynamic, scalable, and in alignment with the CAP theorem, ensuring Consistency, and Partition-Tolerance. Moreover, it aims to incorporate a wide spectrum of EMG features, both in the time and frequency domains. These features are extracted from a diverse participant pool that includes healthy individuals as well as those with spinal cord injuries and stroke. Additionally, the DB integrates anthropometric variables such as gender, age, and Body Mass Index (BMI), with the participant selection based on Greece's average demographic statistics and specific inclusion and exclusion criteria from existing clinical trials from the NeuroSuitUp and Heroes projects.[12,13] The DB not only offers normative sEMG values but also hosts raw sEMG signals, enhancing its utility for researchers and practitioners alike.

METHODS AND MATERIALS

Database Design & Schema

The staggering volume of data generated and processed every day is a defining feature of the modern era. Particularly in biomedical research, where data serve as both input and feedback for useful insights, this "data deluge" poses a unique mix of obstacles and opportunities. The management of such enormous datasets requires DBs that are not only robust but also flexible enough to deal with the peculiarities of the data they are designed to handle. To handle divergent storage challenges, different DBs must be designed depending on the situation.[14,15]

Traditional SQL DBs are well-suited for handling structured data and offer robust query capabilities. These DBs have excelled in storage efficiency and data retrieval speeds when compared to rudimentary flat file systems. However, they often fall short when tasked with handling non-uniform or unstructured data, a characteristic that is increasingly prevalent in today's data-rich environment. NoSQL DBs have emerged as the tool of choice for managing these extensive, heterogeneous, and ever-evolving data sets, leading to the rise of NoSQL DBs, particularly those of a document-oriented nature.[14–16] Studies indicate that NoSQL architectures outperform their SQL counterparts in almost all performance metrics, including the speed of data storage, indexing, and query retrieval.[16]

In this study, MongoDB—a document-based NoSQL DB—is selected as it is well-equipped to manage such inconsistencies. MongoDB resides on the CP side of the CAP theorem (Figure 1), meaning it prioritizes consistency and partition-tolerance over availability.

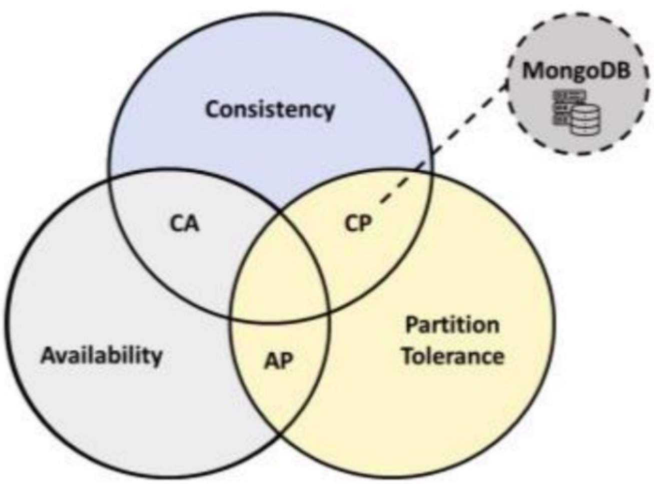

FIGURE 1. MongoDB—CAP theorem.

Though NoSQL DBs like MongoDB are typically schema-less, this study utilizes a conceptual schema to clarify its architectural structure. The schema (Figure 2) consists of multiple tables and each table possesses unique primary keys and defined attributes suitable for storing a variety of data types. The architecture allows for straightforward referencing between tables. This interconnection enhances the DB's robustness, making it adaptable to different query requirements and enhances its scalability to manage dynamic and multi-dimensional data. Moreover, the incorporation of JSON-like documents with dynamic schemas not only simplifies data integration but also offers greater flexibility, thus exemplifying modern DB requirements.

FIGURE 2. Database schema.

Sample Size

In this study, stratified sampling is employed. The population is going to be divided into homogeneous subpopulations, or strata, based on gender.[17] Analyses have suggested that a minimum of 50 subjects per stratum is needed to ensure clinically useful confidence intervals; therefore, a sample size of 100 participants is selected, with 50 participants in each gender-based cell. Previous studies have shown that the inclusion of fewer than 50 subjects per stratum results in confidence intervals that lack clinical utility, while studies with more than 75 subjects per group do not significantly refine these intervals.[18]

The focus on gender as a stratifying variable was predicated on previous research indicating significant gender differences in patterns of muscle fatigue and neuromuscular activation during isometric contractions.[17] This supported the inclusion of gender as a critical stratifying variable to gain insights into muscle fatigability and endurance capacity, which are influenced differently in men and women, in the pursuit of developing a DB with normative data.[17]

It is worth mentioning that the selected sample size takes also into consideration the influence of data skewness on measures of central tendency and variance. For normally distributed data, a stable measure can be obtained with a sample size of approximately 70; however, the requirement may vary between 30 and 80 depending on the skewness of the data. Provision of skewness as a descriptive statistic is highly recommended for future studies, in order to enable clinicians to make more informed decisions.[19]

Database Validation

A multi-step approach is used to validate the DB, ensuring its robustness and clinical applicability. of the data stored. Initially, data are acquired from the main cohort of healthy participants. Following this, sEMG processing and feature extraction techniques are applied to the collected data, leading to time and frequency domain analyses. Subsequently, statistical metrics such as Means, Standard Deviations, and Skewness are calculated. Transforms are applied to these metrics to approximate Gaussian distributions if the initial data deviate from a Gaussian pattern. Z-scores for each subject are then computed. Leave-one-out Gaussian Validation is executed to ensure optimum sensitivity in the Gaussian cross-validation. This method is chosen for its efficacy, despite being less rigorous than completely independent cross-validation, which is often more resource-intensive.[20] Clinical correlations and validity are conducted with a second cohort comprising patients with spinal cord injuries (SCI) or stroke conditions, evaluated by experienced clinicians. Parametric and Non-parametric statistical methods are applied throughout the validation process (Figure 3). The feedback mechanisms between Gaussian cross-validation and statistical metrics, as well as between clinical validation and sEMG processing and feature extraction, are used to fine-tune the DB's performance and relevance, closely aligning it with clinical requirements.[21]

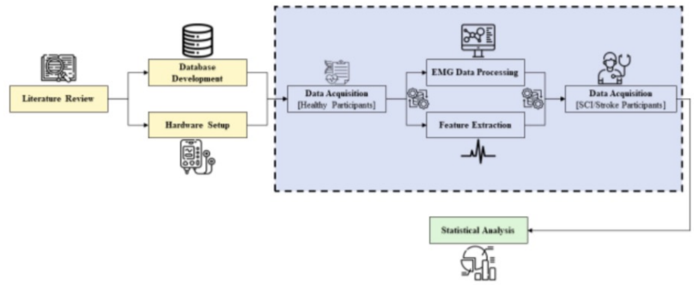

FIGURE 3. Database validation.

FUTURE PERSPECTIVES

Workflow

An integrated methodology is suggested to strengthen the reliability and clinical relevance of our normative EMG DB in advance of future research trajectories. The workflow starts with a per-person profile that considers anthropometric data such as laterality, gender, and age. The sEMG MyoWare 2.0 Muscle Sensor (Advancer Technologies, LLC, version 2.0, Raleigh, NC, USA) is used for sEMG Signal Acquisition to record RAW sEMG data, which is then stored in the DB. A two-tiered computational analysis follows; for the main cohort, an EMG processing and feature extraction process is executed, and the resulting data are then stored in the DB. It's crucial to note that a validation cohort is employed for clinical correlations, as validated by experienced clinicians. These correlations serve to ascertain the relevance of the DB in practical, clinical settings. The validation process for the DB, as outlined in the previous section, follows as the next step of the workflow adding an extra layer of credibility. This workflow, illustrated in Figure 4, seeks to provide a normative sEMG DB.

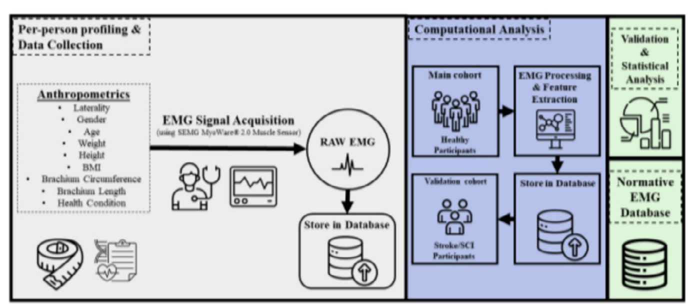

FIGURE 4. Study workflow.

DISCUSSION

The current study addressed the lack of availability of normative DBs for QEMG by designing a NoSQL DB for normative sEMG values of upper body muscles. The absence of such DBs poses a significant obstacle in fields such as physiotherapy and neurorehabilitation, thereby hindering precise diagnostics and effective treatment strategies. In alignment with the CAP theorem, MongoDB is chosen for its ability to manage large and heterogeneous data and a conceptual schema is implemented to clarify the architectural structure and guide the data storage and retrieval processes. The DB's applicability is enhanced by a stratified sampling method focused on gender. A multi-step validation approach is undertaken to ensure the DB's clinical applicability.

However, the study has its limitations, primarily the focus on upper body muscles and the restricted sample size. Although the DB is designed to handle a diverse range of sEMG data, it has the potential to be more comprehensive. Furthermore, stratification can be further expanded, contingent upon an increase in sample size, offering opportunities for future refinement and increasing the level of representation. It is disconcertingly revealed through the study that a contemporary, universal methodology for generating a reliable normative EMG DB is lacking, and that frameworks for QEMG Normative DBs are similarly deficient.

CONCLUSION

This study addresses the notable absence of normative DBs neurorehabilitation research. Capitalizing on the flexibility and scalability of MongoDB, a NoSQL DB is developed, which integrates sEMG data and anthropometric variables from a demographically representative participant pool in Greece, including both healthy participants and post stroke or spinal cord injury patients. The DB is designed to be scalable, enhancing its long-term utility for clinical diagnostics and rehabilitation research. Validation protocols and clinical correlations further refine its practical relevance. The proposed DB can stand as an asset for researchers and is expected to be applied to clinical practice in future studies.

ACKNOWLEDGEMENT

This study was conducted as a part of the research program "NeuroSuitUp: Neurorehabilitation through synergistic man-machine interfaces promoting dormant neuroplasticity in spinal cord injury" (MIS 5047840) and will be implemented and clinically validated in the context of the project "HEROES: Human Extremity Robotic Rehabilitation and Outcome Enhancement for Stroke" funded by H.F.R.I.

CONFLICTS OF INTEREST

The authors declare they have no competing interests

REFERENCES

1. Escabí, M. Biosignal Processing. In *Introduction to Biomedical Engineering*, 3rd ed.; Enderle, J.D. and Bronzino, J.D; Elsevier Science & Technology, Connecticut, USA; 2011; pp. 667–746.
2. Konrad, P. The ABC of EMG: A Practical Introduction to Kinesiological Electromyography, 2006. Available online: https://www.studeersnel.nl/nl/document/universiteit-van-amsterdam/neurologie/the-abc-of-emg-by-peter-konrad/59807737.
3. Electromyography: Physiology, Engineering, and Non-Invasive Applications, 2023. Available online: https://doi.org/10.1002/0471678384.fmatter.
4. Pitt, M.C. and Jabre, J. The problem of lack of normative data in paediatric EMG and possible solutions. *Clin Neurophysiol*. 2018;129(3):672–675. https://doi.org/10.1016/j.clinph.2017.11.024.
5. De Luca, C.J. The Use of Surface Electromyography in Biomechanics. *J Appl Biomech*.1997;13(2):135–163. https://doi.org/10.1123/jab.13.2.135.
6. Farina, D., Merletti, R., Enoka, R.M. The extraction of neural strategies from the surface EMG. *J Appl Physiol*. 2004;96(4):1486–1495. https://doi.org/10.1152/japplphysiol.01070.2003.

7. Meekins, G.D., So, Y., Quan, D. American Association of Neuromuscular & Electrodiagnostic Medicine evidence-based review: Use of surface electromyography in the diagnosis and study of neuromuscular disorders. *Muscle Nerve*. 2008;38(4):1219–1224. https://doi.org/10.1002/mus.21055.
8. Jeppesen, T.D., Levison, L., Codeluppi, L., et al. Quantitative electromyography: Normative data in paraspinal muscles. *Muscle Nerve*. 2020;62(3):358–362. https://doi.org/10.1002/mus.27000.
9. Ryan, C.S., Conlee, E.M., Sharma, R., et al. Nerve conduction normal values for electrodiagnosis in pediatric patients. *Muscle Nerve*. 2019;60(2):155–160. https://doi.org/10.1002/mus.26499.
10. Brewer, E. CAP twelve years later: How the "rules" have changed. *Computer*. 2012;45(2):23–29. https://doi.org/10.1109/MC.2012.37.
11. Davoudian, A., Chen, L., Liu, M. A Survey on NoSQL Stores. *ACM Comput Surv*. 2018;51(2):1–43. https://doi.org/10.1145/3158661.
12. Athanasiou, A., Mitsopoulos, K., Praftsiotis, A., et al. Neurorehabilitation Through Synergistic Man-Machine Interfaces Promoting Dormant Neuroplasticity in Spinal Cord Injury: Protocol for a Nonrandomized Controlled Trial. *JMIR Res Protoc*. 2022;11(9):e41152. https://doi.org/10.2196/41152.
13. Mitsopoulos, K., Fiska, V., Tagaras, K., et al. NeuroSuitUp: System Architecture and Validation of a Motor Rehabilitation Wearable Robotics and Serious Game Platform. *Sensors*. 2023;23(6):3281–3281. https://doi.org/10.3390/s23063281.
14. Rautmare, S. and Bhalerao, D.M. MySQL and NoSQL database comparison for IoT application. In *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*; IEEE; 2016; pp.235–238. https://doi.org/10.1109/ICACA.2016.7887957.
15. Gupta, A., Tyagi, S., Panwar, N., et al. NoSQL databases: Critical analysis and comparison. In *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, Gurgaon, India; IEEE; 2017; pp. 293–299. https://doi.org/10.1109/IC3TSN.2017.8284494.
16. Schulz, W.L., Nelson, B.G., Felker, D.K., et al. Evaluation of relational and NoSQL database architectures to manage genomic annotations. *J Biomed Inform*. 2016;64:288–295. https://doi.org/10.1016/j.jbi.2016.10.015.
17. Clark, B.C., Manini, T.M., Thé, D.J., et al. Gender differences in skeletal muscle fatigability are related to contraction type and EMG spectral compression. *J Appl Physiol*. 2003;94(6):2263–2272. https://doi.org/10.1152/japplphysiol.00926.2002.
18. Bridges, A.J. and Holler, K.A. How Many is Enough? Determining Optimal Sample Sizes for Normative Studies in Pediatric Neuropsychology. *Child Neuropsychol*. 2007;13(6):528–538. https://doi.org/10.1080/09297040701233875.
19. Piovesana, A. and Senior, G. How Small Is Big: Sample Size and Skewness. *Assessment*. 2016;25(6):793–800. https://doi.org/10.1177/1073191116669784.
20. Wu, Y.C. and Lin, I.M. The Resting State of Taiwan EEG Normative Database: Z-Scores of Patients with Major Depressive Disorder as the Cross-Validation. *Brain Sci*. 2023;13(2):351. https://doi.org/10.3390/brainsci13020351.
21. Thatcher, R.W., Walker, R.A., Biver, C.J., et al. Quantitative EEG Normative Databases: Validation and Clinical Correlation. *J Neurother*. 2003;7(3–4):87–121. https://doi.org/10.1300/j184v07n03_05.