Implementation of Subject Specific Model and General Model using Bayesian Algorithm and Compare their Performance

Love Jain
ECE Department
Delhi Technological University
lovejain_ec20b13_02@dtu.ac.in

Nishant Shaklan
ECE Department
Delhi Technological University
nishantshaklan_ec20a14_23@dtu.ac.in

Hemant Kumar
ECE Department
Delhi Technological University
hemantkumar_ec20b12_63@dtu.ac.in

Rajesh Birok
ECE Department
Delhi Technological University
rbirok@dtu.ac.in

***Abstract*—Electromyography (EMG) signals play a pivotal role in numerous applications, offering insights into muscle activity crucial for various fields. However, the challenge lies in mitigating noise interference during transmission, which significantly impacts signal accuracy and reliability. To address this pressing issue, our research delves into the efficacy of subject-specific and general-specific models for EMG signal detection, employing sophisticated Bayesian Inference algorithms.**

This comprehensive investigation encompasses a diverse dataset representing various demographics, including healthy individuals, older adults, and individuals with specific medical conditions. By leveraging this rich dataset, we aim to conduct a robust evaluation of the proposed models, shedding light on their performance across different population segments and scenarios.

The overarching goal of this research is to advance the field of EMG signal detection, with profound implications for neuromuscular diagnosis, prosthetic control systems, and the development of intuitive human-computer interfaces. By enhancing our understanding of EMG signal processing and modeling, we strive to pave the way for innovative solutions that empower individuals with improved healthcare diagnostics and assistive technologies, ultimately enhancing their quality of life.

Keywords: Electromyography (EMG), subject-specific, general-specific, diverse dataset, Bayesian Inference.

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Electromyography (EMG) is an important technique for assessing skeletal muscle electrical activity, utilizing an electromyograph to generate electromyograms. EMG detects electric potentials from activated muscle cells, with frequencies ranging from 0.1Hz to 1kHz [1]. Signal amplitudes range from 0 to 10 microvolts, amplified to 0 to 10 millivolts. Analysis of EM signals provides insights into abnormalities, muscle activation levels, recruitment order, and biomechanical aspects of human or animal movement [2]. Needle EMG, common in neurology, differs from surface EMG, used by professionals like physiotherapists and biomedical engineers [3]. Electromyograms capture muscle activity nuances during rest and contraction, aiding in nerve system exploration and nerve signal transmission speed measurement [4].

This paper examines two primary types of EMG models: Subject-specific methods [5] and General Methods [6]. Subject-specific methods are tailored approaches within specific fields, employing domain-specific knowledge and tools to address unique challenges. They encompass experimental protocols, data collection procedures, analytical frameworks, and computational algorithms, enabling focused investigation of subject-specific phenomena. Conversely, General Methods are systematic approaches applied across disciplines, characterized by adaptability and applicability to diverse scenarios. They guide the analysis, experimentation, or decision-making process with structured steps such as observation, hypothesis formulation, experimentation, data collection, analysis, and conclusion drawing. By comparing these models, this research aims to elucidate their respective strengths and limitations, contributing to a deeper understanding of EMG signal analysis and advancing knowledge in biomedical engineering and related fields.

Here, we utilize machine learning (ML) as the exclusive methodology for our EMG models, motivated by its unmatched capabilities in handling the intricate data patterns inherent in electromyography signals. ML algorithms are celebrated for their proficiency in extracting meaningful insights from vast datasets, facilitating precise classification and interpretation of muscle activity. Our decision to rely solely on ML is rooted in its adaptability and scalability, enabling accommodation of diverse demographic factors such as age, gender, and medical conditions within our analytical framework. By leveraging ML techniques, our goal is to develop robust predictive models capable of identifying abnormalities, tracking muscle activation patterns, and elucidating underlying physiological mechanisms [7]. Through our commitment to ML methodologies, we anticipate substantial advancements in EMG signal processing and interpretation, poised to catalyze innovation in fields including healthcare, rehabilitation, and biomechanics.

II. WHY BAYESIAN INFERENCE

In this paper we have compared two different types of EMG models which are developed by ML algorithms. And specifically we use Bayesian Inference [8] as an ML algorithm for implementing the models of EMG. The use of Bayesian Inference in implementation of the EMG models offers several advantages like:

- **Incorporation of Prior Knowledge**: Bayesian inference allows researchers to incorporate prior knowledge or beliefs about the parameters of the model into the analysis. This is particularly useful in EMG modeling, where domain expertise can provide valuable insights into the underlying physiological processes.
- **Handling of Uncertainty**: Bayesian inference provides a framework for quantifying uncertainty in model parameters. This is important in EMG modeling, where parameters such as muscle activation levels or signal noise may be uncertain due to various factors such as electrode placement or biological variability.
- **Regularization and Model Selection**: Bayesian methods naturally lend themselves to regularization techniques, which can help prevent overfitting in EMG models, especially when dealing with high-dimensional data or limited sample sizes. Additionally, Bayesian model comparison methods allow for the comparison of different EMG models, facilitating model selection.
- **Sequential Learning and Updating**: Bayesian inference allows for sequential learning and updating of model parameters as new data becomes available. This is advantageous in EMG applications where data may be collected over time, such as in longitudinal studies or real-time monitoring scenarios.
- **Integration of Hierarchical Models**: Bayesian hierarchical modeling enables the incorporation of hierarchical structures into EMG models, allowing for the modeling of complex dependencies and interactions between different levels of the data. This can be particularly useful when analyzing EMG data collected from multiple subjects or different experimental conditions.
- **Propagation of Uncertainty**: Bayesian inference provides a principled framework for propagating uncertainty from input variables to model predictions. This is important in EMG modeling for accurately assessing the uncertainty associated with predictions of muscle activity or other physiological parameters.

Overall, the use of Bayesian inference in the implementation of EMG models offers a flexible and powerful approach for analyzing EMG data, incorporating prior knowledge, handling uncertainty, and facilitating model selection and inference.

III. SETUP AND SIMULATION

In this research endeavor, we undertook a comprehensive investigation into the efficacy of machine learning algorithms for healthcare analytics using a diverse and extensive dataset. Leveraging the cloud-based computing capabilities of Google Colab, we orchestrated the entire computational process, enabling seamless integration with powerful libraries such as NumPy, Pandas, and Scikit-learn. The choice of Python as the primary programming language facilitated rapid prototyping and efficient implementation of algorithmic solutions. Our methodology encompassed a multifaceted approach, combining Bayesian inference techniques with traditional machine learning paradigms such as SVM [9] and KNN [10].

To ensure the robustness and reliability of our findings, we meticulously designed our experiments, initializing all channel values to zero to preemptively mitigate any potential introduction of spurious data artifacts. The dataset itself comprised an expansive collection of over a million data points, capturing nuanced patient information across various temporal intervals. This rich repository encapsulated a wide spectrum of demographic attributes, spanning diverse age cohorts and encompassing a spectrum of health conditions, meticulously delineated across eight distinct channels.

Our analytical framework prioritized the comprehensive evaluation of model performance, employing a suite of rigorously chosen evaluation metrics including F1 score, precision, recall, and accuracy [11]. By meticulously scrutinizing the variances in these metrics across different algorithms, we gleaned invaluable insights into the nuanced performance nuances of each approach. This granular analysis not only elucidated the comparative strengths and weaknesses of specific algorithms but also shed light on their suitability for both subject-specific and general predictive modeling tasks within the realm of healthcare analytics. Through our meticulous experimentation and rigorous evaluation, we not only contributed to advancing the state-of-the-art in machine learning-driven healthcare analytics but also provided tangible guidance for practitioners and researchers alike, empowering them with actionable insights for informed algorithm selection and model optimization strategies.

IV. METHODOLOGY

This study delves into the effectiveness of Machine Learning (ML) algorithms for analyzing electromyography (EMG) signals, with a specific focus on comparing the performance of **subject-specific and general models**. By leveraging personalized models tailored to individual users, the research aims to assess their capability in capturing **individual variability** within EMG patterns. This personalized approach holds significant promise for optimizing EMG-based simulations and advancing diverse applications across various fields, including prosthetics control, rehabilitation programs, and human-computer interaction systems.

A. Computational Platform

Google Colaboratory (Colab) served as the primary platform for training and evaluating the efficacy of the chosen ML models. This cloud-based environment provided a convenient and accessible infrastructure for conducting these computations.

B. Exploration of Machine Learning Algorithms

The research investigated the suitability of three prominent ML algorithms for EMG analysis: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Bayesian Inference. KNN is a non-parametric supervised learning algorithm that classifies data points based on their similarity to labeled data points in the training set. SVM is another supervised learning algorithm that excels at creating hyperplanes to effectively separate data points belonging to different classes. Bayesian Inference, a probabilistic approach, allows for incorporating prior knowledge and updating beliefs based on new evidence, potentially leading to more robust models.

Baysian

$$\mathrm{P(c|x)} = \frac{\mathrm{P(x|c)P(c)}}{\mathrm{P(x)}} \tag{1}$$

where,

P(c) is the probability of class c

P$(x|c)$ is the conditional probability of x given class c

P(x) is the a prior probability of x

$P(x, c)$ is the joint probability density function of x and c

KNN

$$d_{AB} = \sqrt{\sum_{i=1}^{d}(A_i - B_i)^2} \tag{2}$$

where,

d_{AB} represents the Euclidean distance between points A and B

d is the dimensionality of the feature space (number of features)

A_i represents the i-th feature value of data point A

B_i represents the i-th feature value of data point B

SVM

Consider a binary classification problem with two classes, labeled as +1 and -1. We have a training dataset consisting of input feature vectors X and their corresponding class labels Y.

The equation for the linear hyperplane can be written as:

wx + b = 0

The vector W represents the normal vector to the hyperplane, i.e., the direction perpendicular to the hyperplane. The parameter b in the equation represents the offset or distance of the hyperplane from the origin along the normal vector w.

C. Model Development Strategy

- **Model Architecture Design:** To mimic the real-world EMG acquisition process, a model with eight channels was designed. This number corresponds to the eight electrodes typically employed in EMG recordings, where each electrode captures electrical activity from a specific muscle group [12]. Each channel within the model received a simulated voltage reading, replicating the input data obtained during real EMG measurements. These simulated voltage readings served as the foundation for training the ML models.

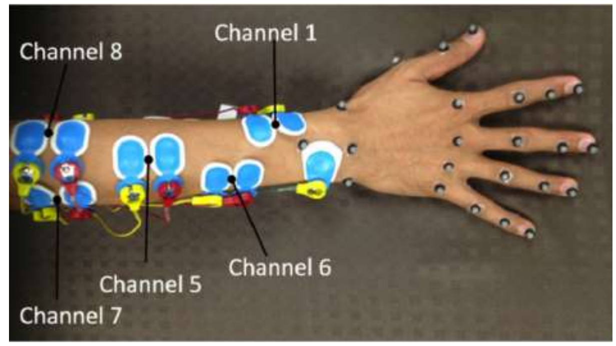

Fig. 1. visual representation of 8 channels used to measure emg signal

- **Development of Model Types:** For each of the explored ML algorithms (KNN, SVM, and Bayesian Inference), two distinct model types were developed:
  - **General Model:** This model was trained on a comprehensive dataset encompassing a wide range of EMG signatures. This dataset was meticulously constructed to incorporate EMG data associated with various activities, muscle groups, age groups, health conditions, and diverse backgrounds. Data acquisition leveraged online platforms like Kaggle to ensure the dataset's richness and diversity, allowing the general model to learn from a broad spectrum of EMG patterns.
  - **Subject-Specific Model:** This personalized model was specifically tailored to individual users by leveraging their unique EMG data. This approach aims to capture the inherent variability in muscle activation patterns that exists between individuals. By focusing on an individual's specific EMG characteristics, the subject-specific model has the potential to achieve a more nuanced understanding of their muscle activity.

D. Data Preprocessing

To ensure compatibility with the chosen ML algorithms and facilitate efficient analysis, all collected data underwent a meticulous conversion process. This preprocessing step involved data cleaning, normalization, and potentially feature scaling techniques to ensure the data adheres to the specific requirements of the chosen algorithms.

E. Model Development Strategy

- **Derived Features (SNS Signals/Ratios)** This approach involved applying specific mathematical operations to the raw EMG data to create new features [13]. Examples include calculating the Signal-to-Noise Ratio (SNR) [14] or other ratios derived from voltage readings across different channels. These derived features aimed to provide additional insights beyond the raw voltage data. Formulas used to calculator these derived feature are:

$$\mathrm{MAV} = \frac{1}{\eta}\sum_{i=1}^{n}|x_i| \tag{3}$$

where,
MAV is the mean absolute value
n = total number of values in the set
x_i = each individual value in the set
$|x_i|$ = absolute value of each x_i

$$\lambda = \frac{v}{f} \tag{4}$$

where,
λ = wavelength of the signal
v = speed of the signal in the medium
f = frequency of the signal

$$\mathrm{Sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0, \\ -1, & x(n) < 0. \end{cases} \tag{5}$$

$$g = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)s^3} \tag{6}$$

where,

- g is the sample skewness
- $x(n)$ is the nth sample value
- $\mathrm{Sgn}[x(n)]$ is the sign function of $x(n)$ * 1 if $x(n) \geq 0$ * -1 if $x(n) < 0$
- x_i is the i-th sample value
- $\bar{x}$ is the sample mean (average)
- n is the total number of samples
- s is the sample standard deviation

$$\mathrm{Kurt} = \frac{\sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s} \right)^4}{(n-1)} \tag{7}$$

where:

- Kurt is the sample kurtosis
- n is the number of observations in the sample
- x_i represents each individual observation
- $\overline{x}$ represents the sample mean (average)
- s represents the sample standard deviation

$$\mathrm{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \tag{8}$$

where:

- RMS is the root mean square value
- n is the total number of values in the set
- x_i represents each individual value in the set

$$\mathrm{SSI} = \int_a^b [f(t)]^2 \, dt \tag{9}$$

where:

- SSI is the Simple Square Integral
- $f(t)$ is the signal as a function of time (t)
- $[a, b]$ represents the interval of integration, where a is the lower bound and b is the upper bound
- $\int_a^b$ denotes the definite integral over the interval $[a, b]$

- **Raw EMG Signals** The model directly received the raw voltage readings from all eight channels. This approach mimicked the real-world EMG acquisition process, where raw voltage data is collected from multiple electrodes placed on the user's skin. By feeding the raw data directly into the model, we investigated the effectiveness of different algorithms and model types in extracting valuable insights directly from the unprocessed EMG signals, potentially reducing reliance on predefined features. Formula to anaylyse confusion matrix thus obtained for bayesian with derived features and bayesian with raw signals:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

V. Limitations

This initial research employed simulated data instead of actual EMG recordings. This initial exploration focused on establishing the fundamental functionalities of the model architecture, evaluating the suitability of different ML algorithms and model types for EMG analysis, and investigating the potential of feature engineering techniques. Future studies will incorporate real-world EMG data for comprehensive performance evaluation and practical application development. By exploring the effectiveness of **subject-specific and general models** for EMG analysis, this research lays the groundwork for advancing personalized approaches to understanding and utilizing EMG signals. This personalized approach has the potential to unlock new possibilities in various fields by capturing the unique nuances of individual muscle activity patterns.

VI. Results and Discussions

In the realm of electromyography (EMG) classification, the adoption of feature processing techniques stands out as a pivotal strategy, demonstrating substantial performance enhancements when compared to the utilization of raw features alone. Through the refinement of input representations, feature processing techniques consistently yield notable improvements of approximately 10% in classification accuracy. This underscores the critical role of feature engineering in optimizing the discriminative power of machine learning models for EMG analysis.

Bayesian models emerge as a robust framework for EMG classification [15], consistently outperforming traditional machine learning approaches across various classification tasks. Leveraging a probabilistic paradigm, Bayesian models excel in handling uncertainty inherent in EMG data and effectively capture complex relationships within the signals. Their superior performance underscores their suitability for addressing the intricate challenges posed by EMG classification tasks.

Notably, subject-specific Bayesian models offer a significant advancement over generalized models in the realm of EMG classification. By training on subject-specific data, these models harness the diverse nuances and variations present in individual subjects' EMG signals. As a result, subject-specific Bayesian models can more accurately capture the unique characteristics of each subject's muscular activity, leading to refined and personalized classification outcomes. This highlights the promise of subject-specific modeling approaches in further enhancing EMG classification accuracy and advancing the field of EMG-based applications.

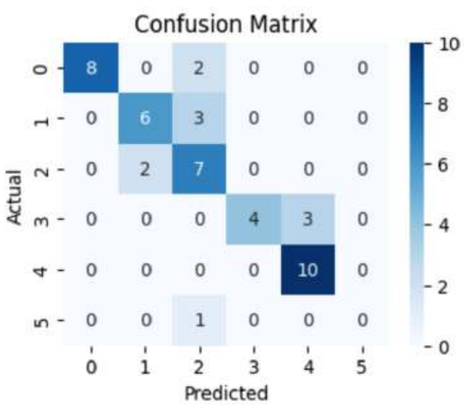

Fig. 2. Confusion Matrix of General Model

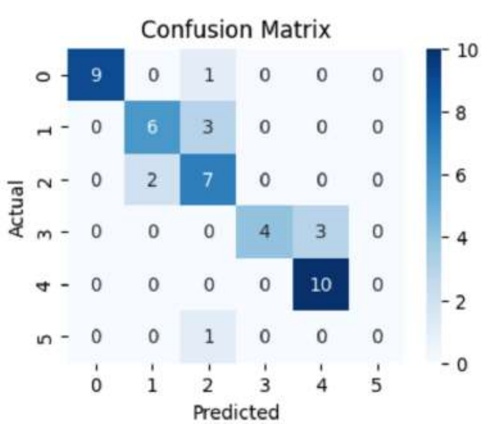

Fig. 3. Confusion Matrix of Subject-Specific Model

A. Model Performance

Tables I and II present the performance metrics of Bayesian and traditional (non-Bayesian) models, respectively, using SNS signals/ratios as features.

TABLE I
PERFORMANCE METRICS OF BAYESIAN MODELS

Model	Accuracy	Precision	Recall	F1 Score
General	0.8688	0.8684	0.8688	0.8500
Subject-Specific	0.9848	0.9848	0.9848	0.9848

TABLE II
PERFORMANCE METRICS OF TRADITIONAL MODELS (NON-BAYESIAN)

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.6232	0.6386	0.6232	0.6145
KNN	0.6223	0.6367	0.6223	0.6131

VII. CONCLUSION

An investigation into machine learning for EMG analysis explored the impact of feature engineering and model selection. The analysis revealed that traditional models, like KNN and SVM, achieved higher accuracy when using raw EMG data (voltage readings) compared to derived features (e.g., signal-to-noise ratios). This suggests that for these algorithms, raw data might be more effective than relying on handcrafted features.

Separately, researchers compared subject-specific and general models using a Bayesian model with both raw and feature-engineered data (e.g., signal-to-noise ratios). The subject-specific model with feature-engineered data outperformed others. This finding highlights the potential benefits of combining Bayesian inference with feature engineering techniques for personalized EMG analysis, offering an advantage over alternative approaches.

REFERENCES

[1] G. Allison and T. Fujiwara, "The relationship between emg median frequency and low frequency band amplitude changes at different levels of muscle capacity," *Clinical Biomechanics*, vol. 17, no. 6, pp. 464–469, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268003302000335

[2] D. F. Stegeman, J. H. Blok, H. J. Hermens, and K. Roeleveld, "Surface emg models: properties and applications," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 313–326, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1050641100000237

[3] "Chapter 16 - needle electromyography: Basic concepts," in *Clinical Neurophysiology: Basis and Technical Aspects*, ser. Handbook of Clinical Neurology, K. H. Levin and P. Chauvel, Eds. Elsevier, 2019, vol. 160, pp. 243–256. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444640321000163

[4] I. Saad, N. H. Bais, C. Bun Seng, H. M. Zuhir, and N. Bolong, "Electromyogram (emg) signal processing analysis for clinical rehabilitation application," in *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, 2015, pp. 105–110.

[5] S. Ma, C. Chen, D. Han, X. Sheng, D. Farina, and X. Zhu, "Subject-specific emg modeling with multiple muscles: A preliminary study," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 740–743.

[6] P. Jayaweera, "Design and implementation of electromyography (emg) based real-time pattern recognition model for prosthetic hand control," 09 2021.

[7] A. Vijayvargiya, P. Singh, R. Kumar, and N. Dey, "Hardware implementation for lower limb surface emg measurement and analysis using explainable ai for activity recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.

[8] H. Han and S. Jo, "Supervised hierarchical bayesian model-based electomyographic control and analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1214–1224, 2014.

[9] M. Z. Al-Faiz, A. A. Ali, and A. H. Miry, "A k-nearest neighbor based algorithm for human arm movements recognition using emg signals," in *2010 1st International Conference on Energy, Power and Control (EPC-IQ)*, 2010, pp. 159–167.

[10] ——, "A k-nearest neighbor based algorithm for human arm movements recognition using emg signals," in *2010 1st International Conference on Energy, Power and Control (EPC-IQ)*, 2010, pp. 159–167.

[11] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, S. Eger, Y. Gao, M. Peyrard, W. Zhao, and E. Hovy, Eds., Nov. 2020.

[12] Y.-M. Wong and G. Y. Ng, "Surface electrode placement affects the emg recordings of the quadriceps muscles," *Physical Therapy in Sport*, vol. 7, no. 3, pp. 122–127, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1466853X06000538

[13] A. Subasi, "Classification of emg signals using combined features and soft computing techniques," *Applied Soft Computing*, vol. 12, no. 8, pp. 2188–2198, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494612001330

[14] J. G. Kreifeldt, "Signal versus noise characteristics of filtered emg used as a control source," *IEEE Transactions on Biomedical Engineering*, vol. BME-18, no. 1, pp. 16–22, 1971.

[15] S. Park, W. K. Chung, and K. Kim, "Training-free bayesian self-adaptive classification for semg pattern recognition including motion transition," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 6, pp. 1775–1786, 2020.