POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Analysis of Neural Activity Changes Associated with LRRK2 Gene Mutation:

A Computational Approach for Neuronal Activity Classification Using In-Vitro Networks

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Author: Gaia Vettori

Advisor: Prof. Riccardo Barbieri

Academic year: 2023-2024

---

1. Introduction

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by motor dysfunction, including tremors, rigidity, bradykinesia, and postural instability. Genetic factors play a significant role in its pathogenesis, with the G2019S mutation in the leucine-rich repeat kinase 2 (LRRK2) gene identified as a common contributor to both familial and sporadic forms of PD. Despite the established association between the LRRK2 G2019S mutation and PD susceptibility, the precise mechanisms underlying neurodegeneration remain unclear. Recent advancements in neurotechnology, particularly microelectrode array (MEA) systems, offer new avenues for studying neuronal activity with unprecedented precision.
This thesis utilizes MEA technology and computational methodologies to investigate the impact of the LRRK2 G2019S mutation on neuronal activity. By developing a comprehensive computational pipeline, structured into spike sorting, point process modeling, and classification, we aim at identifying distinct signatures of neural dysfunction in LRRK2-mutated neuronal networks, providing insights into PD pathogenesis.

2. Background

2.1. Neuron signals

The understanding of neurons, essential for unraveling brain complexities, reveals their structure as an input-output system — comprising a cell body, dendrites, and an axon — where dendrites receive signals and axons transmit them. A study in 1994 [4], demonstrates action potential initiation in the axon, followed by propagation to dendrites, supported by experiments on cerebellar Purkinje cells showing decreasing amplitude with distance from the soma. Studies also reveal distinct electrical components in neuronal recordings, including synaptic activity, local field potentials, inhibitory currents, and return currents. From these findings, we learn that action potentials are generated by positive charge that appear as a negative difference from external recordings; on the other hand, return currents flowing through the dendrites, being negatively charged, will appear as a positive peak. This distinction is important to define the spike detection and sorting algorithm later on.

2.2. Neuron modeling

The Integrate-and-Fire model provides a mathematical description of the dynamics of the neuronal membrane potential and action potential generation. Compared to the other common alternative model, the Hodgkin and Huxley (HH) model, the IF model offers a simplified yet effective framework for under-

standing the neuron's behavior. Described by the equation:

$$\mathrm{d}V_t = -L(V_t - V_{\text{rest}})\mathrm{d}t + dI_{\text{syn}}(t) \tag{1}$$

the evolution of neuronal membrane potential over time incorporates excitatory and inhibitory components in the synaptic current input, with a spike generated when the membrane potential reaches a threshold. In IF models, when that happens, the membrane potential is reset to its resting state, contrarily to HH models, which provide a more detailed representation of membrane dynamics throughout the action potential process. Furthermore, when dealing with Integrate-and- Fire models, it has been proven that the firing activity is modeled as a Wiener process, following a Brownian motion.

$$V(t) = V_0 + \beta t + \int_0^t dW(u)du \tag{2}$$

And its Probability Density Function follows an Inverse Gaussian Distribution [1].

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) \tag{3}$$

2.3. Induced Pluripotent Stem Cells

Pluripotency, traditionally defined as the capability to generate most cell types of an embryo, is embodied by embryonic stem (ES) cells, derived from mammalian blastocysts' inner cell mass. These cells, with indefinite growth and pluripotency maintenance, hold promise for disease understanding, drug screening, and therapeutic applications. However, ethical concerns and challenges in generating patient- or disease-specific ES cells impede their widespread use.

Direct reprogramming of somatic cells to induce pluripotency offers a solution. The method essentially involves reprogramming adult human cells to behave like ES cells, but without using human embryos. The breakthrough discovery in 2007 by Takahashi and Yamanaka demonstrated the feasibility of this approach [5]. By using retroviral transduction and optimized culture conditions, adult cells were successfully converted into a pluripotent state, offering a promising alternative for regenerative medicine and personalized therapies without ethical concerns related to human embryo use.

2.4. Parkinson's Disease and LRRK2-mutation

Parkinson's Disease is a neurological disorder mainly characterized by motor impairments. Its prevalence was around 1 in 1000 people by the end of the last millennium, primarily affecting individuals over 50 years old. The disease results from the deterioration of the neuronal cytoskeleton, particularly affecting cells in the limbic and motor systems, leading to the formation of Lewy bodies and neurites, abnormal protein aggregates that accelerate neuronal degeneration. Diagnosis typically relies on visible symptoms, confirmed post-mortem. Early Parkinson's symptoms can be diverse and nonspecific, including tremors, sensory disturbances, muscle pain or weakness, difficulty with tasks like handwriting, fatigue, depression, and unexplained weight loss. Due to this variability, initial diagnosis may be challenging, with symptoms often attributed to other medical conditions.

The leucine-rich repeat kinase 2 (LRRK2) protein, with GTPase and kinase domains, is linked to Parkinson's disease (PD), particularly through the G2019S mutation, present in 4% of familial and 1% of sporadic cases worldwide [2]. PD models suggest LRRK2 mutations affect various cellular processes, including vesicular trafficking and protein synthesis. Clinical and pathological analysis of PD patient populations have shown the causal relationship between LRRK2 kinase domain mutations and familial PD onset. A study examining the G2019S mutation prevalence among PD patients found 5% carried the mutation, with carriers showing longer disease duration and increased *levodopa* use [3]. Mechanistically, LRRK2 or *Dardarin*, crucial in the process of protein generation, vital for neuronal health and functions, suggests that alterations in LRRK2 can increase neuronal susceptibility to degeneration and protein aggregation. These findings underscore the significance of LRRK2 mutations in PD susceptibility, potentially informing genetic testing for the disease.

3. The Data

The data utilized in this study comprises recordings from 12 neural networks of cortical in-vitro neurons, half of which carry the genetic mutation of the LRRK2 gene. To induce pluripotency, Episomal iPSC reprogramming vectors are employed on dermal fibroblasts using Yamanaka factors. Subsequently, the cells are cultured and expanded on poly-L-Ornithine and laminin, followed by placement in a standard humidified air incubator for neural stem cell growth. Synchronized differentiation into cortical neurons occurs over 15 days using NSC reagent bundles and media. Recordings are obtained from resulting cultures, starting from the baseline recordings, circa 400 seconds long. Half of each neuron population undergo chemical stimulation with Kainic Acid, applied for 30 minutes, followed by recordings 24 hours post-stimulation. The chemical component is meant to provoke neuro-toxicity effects. The electrical activities were recorded through the use of Multi-Electrode Array (MEA) technology with 3-nodal microfluidic chips, specifically cus-

tomized by Van de Wijdeven et al [6]. Each nodal chamber, with an octagonal shape of 4mm in diameter, accommodates neurons for growth and development. The chambers are interconnected via microtunnels (104 in total), including axons and dendrites from neighboring neurons. Beneath the microfluidic chips, 59 micro-electrodes record the neuronal activity, alongside a reference electrode. The resulting files are converted to .h5 format for computational processing.

4. Spike Sorting

The spike sorting process consists of processing raw MEA data to extract spikes and assign them to individual neurons, thereby defining each neuron's complete activity.
Initially, the signal undergoes filtering with a bandpass Butterworth filter ranging from 300 to 3000Hz, followed by the application of a notch filter at critical frequencies on the Reference channel, which is subsequently subtracted from every other channel. To discern spikes from background noise, a threshold is set at 3 times the Median Absolute Deviation (MAD) of the signal over one-minute windows. After initial peak detection, each spike is extracted as a 3ms window around the peak index and evaluated to determine if it meets the criteria for a real spike. These criteria primarily involve mean and standard deviation considerations, as noisy spikes tend to exhibit high values in these metrics. Once the spikes are selected, they undergo standardization and the first three Principal Components, obtained through PCA, are clustered using the k-means clustering method. The cluster composition with the highest silhouette score is chosen. If no cluster configuration achieves a silhouette score above 0.5 (the standard threshold), a single cluster is defined.

5. Point Process Modeling

Point Process Modeling is a method used to characterize the inter-spike intervals (ISIs) of neurons, reflecting their firing activity characteristics. In ideal Integrate-and-Fire models, neuronal firing activity follows a Brownian motion pattern, making the inverse Gaussian distribution the most suitable for modeling the neurons' ISIs. In our study, which examines both healthy neuron populations and those carrying a genetic mutation, the Dirichlet Mixture model emerged as the most effective. This model comprises one inverse Gaussian and two Gaussian distributions. The inverse Gaussian and the first Gaussian distributions are focused on lower frequency values to capture the main peak of ISIs, while the second Gaussian distribution accounts for the tail. The ranges of the parameters (prior probability) are set as follows:

- Inverse Gaussian mean (μ_1): 0-0.2
- Inverse Gaussian lambda (λ): 0.0001-0.1
- 1st Gaussian mean (μ_2): 0-0.2
- 1st Gaussian sigma (σ_2): 0.01-0.7
- 2nd Gaussian mean (μ_3): 0.2-0.6
- 2nd Gaussian sigma (σ_3): 0.01-0.7

The model for each neuron is chosen using Maximum A Posteriori estimation, based on Bayes' theorem. Bayes' theorem, denoted as:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \tag{4}$$

represents the relationship between the posterior probability $P(\theta|y)$, the likelihood function $P(y|\theta)$, and the prior probability $P(\theta)$ of the parameters θ. In our context, we define the prior probabilities of parameters based on their predetermined ranges, as listed above, while the likelihood function is derived from the inter-spike intervals data. By employing MAP estimation, we aim to identify the parameter values that maximize the posterior probability distribution $P(\theta|y)$. Each model's performance is then evaluated using KS distance metrics and KS plots.

6. Classification

The final stage of our processing pipeline encompasses classification tasks, which include a binary classification and a 4-class problem. In the binary classification, networks are categorized into control and LRRK2-mutated populations. Meanwhile, the 4-class problem distinguishes between control vs. mutated networks and baseline vs. post-stimulation activity. Transitioning from binary to multi-class classification impacts certain classifiers differently. While decision tree and random forest classifiers seamlessly adapt to the new framework, logistic regression and support vector machine classifiers utilize strategies such as one vs. rest or one vs. all. Through our evaluation process, we identified the most effective classifiers, by evaluating several performance metrics and choosing the one with the highest scores. When two classifiers both had similarly high performances, the one with the highest average score, computed by averaging all the metrics, was selected.

7. Results

7.1. Spike Sorting

In the spike sorting analysis, significant differences emerge in the firing rates between healthy and genetically mutated populations. Control networks exhibit a mean firing rate of 23Hz, contrasting with the mutated networks' mean firing rate of 6Hz (Figure 1).
Notably, post-stimulation control populations display increased activity, while stimulated mutated networks show a slight decrease. A total of 464

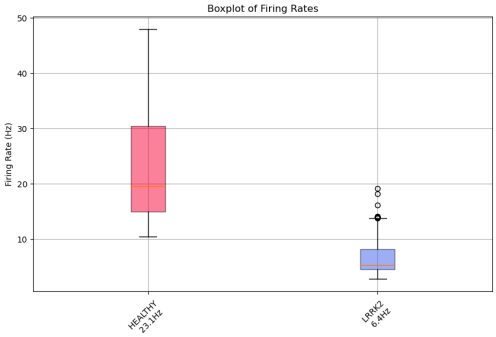

Figure 1: Boxplots of Firing Rates.

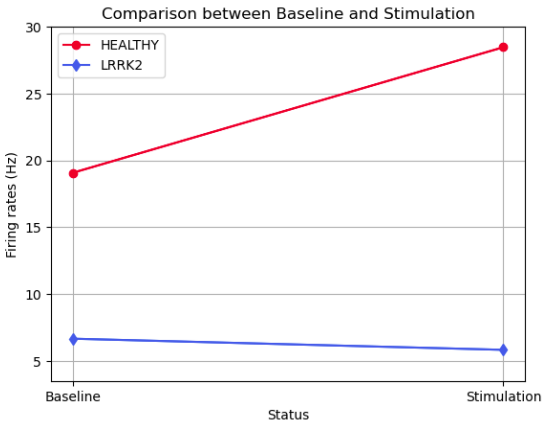

Figure 2: Evolution of Firing Rates.

control neurons (232 pre-stimulation, 232 post-stimulation) and 543 LRRK2-mutated neurons (377 pre-stimulation, 166 post- stimulation) were identified with the Spike Sorting procedure.

7.2. Point Process Modeling

The obtained neurons are then modeled with the Dirichlet Mixture Model composed of the 3 distributions, defined by the parameters in the pre-set ranges, to be the optimal descriptors of the data. The models are evaluated with the KS score distance, which computes the vertical distances between the to-be-modeled data and the chosen model. By analyzing the distribution of the KS distance values, the following conclusions are drawn. The KS distance is on average > 0.15 when the neuron carries the genetic mutation of LRRK2 and is < 0.15 when the neuron is healthy. This means that the model performs better when the neuron belongs to the healthy population. Upon closer inspection, it is noteworthy to observe the contrasting phenomena following chemical stimulation in the two types of populations. As it is illustrated in figure 3, the model performs better on stimulated control neurons rather than on their baseline activity. On the contrary, the KS distances increase in value when the LRRK2-mutated neurons have undergone stimulation, compared to before they have.

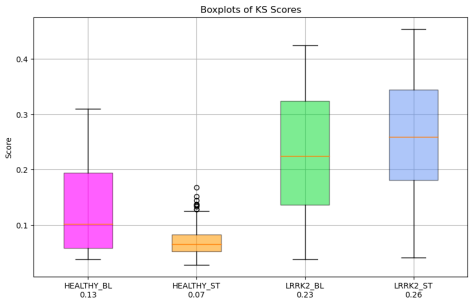

Figure 3: Boxplots of KS distances.

7.3. Classification

Once the ISI distributions' parameters are estimated, the classification tasks consist of two parts:

- Binary Classification: This involves distinguishing neurons from LRRK2-mutated populations and control networks.
- Multi-Class Classification: Neurons are classified into four categories based on population type and activity status (control networks vs. LRRK2-mutated networks and baseline activity vs. post-stimulation activity).

In the binary problem, all classifiers performed exceptionally well, with an accuracy and ROC AUC exceeding 97%, with the Support Vector Machine achieving the highest metrics — reaching 99% accuracy and ROC curve of 99%. An analysis of feature importance identifies μ_1 — the mean of the Inverse Gaussian distribution — as the most influential parameter. Control populations exhibit lower μ_1 values compared to LRRK2-mutated populations, indicating that the Inverse Gaussian distribution for control networks is centered at low frequencies (Figure 4).

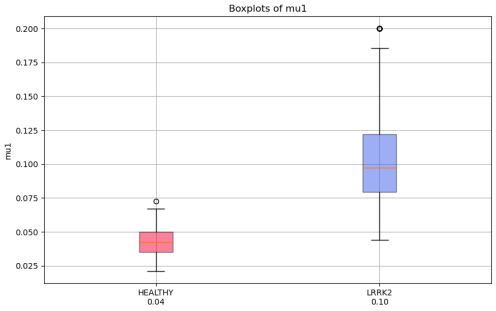

Figure 4: Boxplots of μ_1.

Additionally, there are noticeable differences in the ranges of w_1, representing the weight assigned to the IG distribution (Figure 5). Control networks are predominantly modeled by a single Inverse Gaussian distribution, while LRRK2-mutated networks display less influence from the IG distribution, indicating disparate modeling of characteristic peaks. The

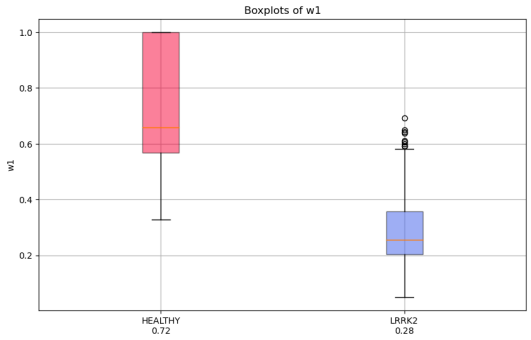

Figure 5: Boxplots of w_1.

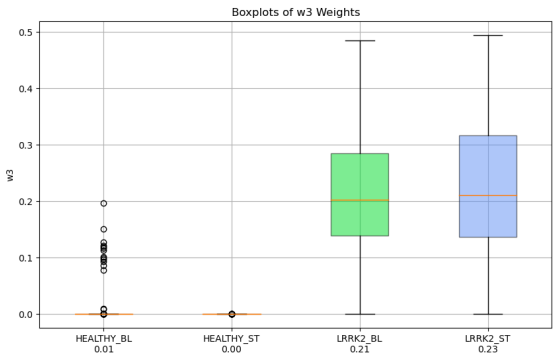

Figure 7: Boxplots of w_3.

table below illustrates how control networks have a predominant IG component (with the weight w_1) and an almost non-existing 2nd Gaussian (w_3); and LRRK2-mutated networks have the main descriptor as the 1st Gaussian curve (w_2) and, in equal parts, the other two distributions (w_1, w_3) (Table 1).

	Healthy	**LRRK2**
w_1	**0.72**	0.279
w_2	0.277	**0.5**
w_3	0.003	0.221

Table 1: Model weights.

The multi-class classification model achieves an accuracy of 76%, with the primary challenge lying in distinguishing between baseline and post-stimulation activities within the same populations. This challenge is clearly observable in the following graph, where the curves obtained by using the mean values of every parameter for every class, are displayed (Figure 6). Feature analysis of the Random Forest

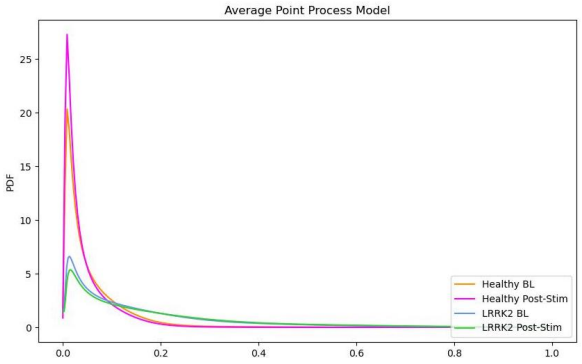

Figure 6: Mean-value Distributions.

classifier highlights w_3, μ_1, and w_1 as the most influential features, with w_3 showing minimal differences between activity phases within the same population type (Figure 7).
A similar phenomenon is observable in the variables μ_1 and w_1 that exhibit significant difference in distribution between control and mutated networks, but not as much between activity status within the same population. This framework poses the greatest challenge for the classifiers that, as a result, can't reach optimal performances.

8. Conclusions

The key points obtained in the discussed results are:

- The control population closely aligns with an almost "pure" Inverse Gaussian model (Table 1), consistent with the theoretical framework outlined in *Section 2.2*.
- Post-stimulation activity demonstrates divergent trajectories in the two populations (Figure 2), suggesting a potential disruption in neuron functioning due to the genetic mutation.

The study presented builds upon existing neuronal modeling work, aiming to refine the existing framework by integrating signal processing techniques, Bayesian classification, and machine learning methods into a comprehensive pipeline for processing, modeling, and classifying raw MEA data into 2 and 4 classes. While the focus has been on distinguishing control versus G2019S LRRK2 genetic mutation samples, the pipeline has broader applicability. Key improvements and advancements employed in this work include enhancing the Spike Sorting algorithm through targeted refinements, improving Point Process Modeling by fine-tuning parameter distributions, and refining the Classification phase outcomes. Although our comprehensive pipeline yielded promising results, areas for enhancement remain, particularly in spike sorting algorithms. Opportunities for refinement exist in exploring alternative classification scenarios beyond the current focus of control versus LRRK2-mutated samples.
This study leveraged advanced neurotechnology and computational methodologies to dissect the underlying mechanisms of neurodegeneration associated with the G2019S mutation in the LRRK2 gene. Through characterizing aberrant neuronal activity patterns, we shed light on potential pathophysio-

logical pathways implicated in Parkinson's Disease progression.
The reported differences between the healthy population and the mutated one, correlate with expected physiological considerations. As reported in *Section 2.4*, the LRRK2 gene is strictly connected to the neuron's well-being and functional behaviors, and its mutation predisposes it to degeneration and malfunctions, observed especially after the KA stimulation, which is meant to provoke neurotoxicity. While achieving its primary objective of sorting, modeling, and classifying, with excellent performance results and providing a complete computational pipeline for neuronal data analysis, this work also contributes to a deeper understanding of neuronal dynamics at a physiological level, crucial in the field of neurodegenerative disorders such as Parkinson's Disease.

References

[1] Riccardo Barbieri, Michael C Quirk, Loren M Frank, Matthew A Wilson, and Emery N Brown. Construction and analysis of non-poisson stimulus-response models of neural spiking activity. *Journal of neuroscience methods*, 105(1):25–37, 2001.

[2] M Ian, Jungwoo Wren, and D Valina. Lrrk2 pathology in parkinson's disease. *Journal of Neurochemistry*, 131:554–565, 2014.

[3] W C Nichols, N Pankratz, D Hernandez, C paisan Ruiz, and S Jain. Genetic screening for a single common lrrk2 mutation in familial parkinson's disease. *Lancet*, 365:410–412, 2005.

[4] G Stuart and M Hausser. Initiation and spread of sodium action potentials in cerebellar purkinje cells. *Neuron*, 13:703–712, 1994.

[5] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872, 2007.

[6] Vibeke Devold Valderhaug, Ola Huse Ramstad, Rosanne van de Wijdeven, Kristine Heiney, Stefano Nichele, Axel Sandvig, and Ioanna Sandvig. Structural and functional alterations associated with the lrrk2 g2019s mutation revealed in structured human neural networks. *BioRxiv*, pages 2020–05, 2020.