Machine Learning to Diagnose Neurodegenerative Multiple Sclerosis Disease

Jin Si Lam[1], Md Rakibul Hasan[2][0000−0003−2565−5321], Khandaker Asif Ahmed[3][0000−0003−1271−7781], and Md Zakir Hossain[1,3][0000−0003−1892−831X]

[1] Australian National University, Canberra, Australia
{jinsi.lam,zakir.hossain}@anu.edu.au
[2] BRAC University, Dhaka, Bangladesh
rakibul.hasan@bracu.ac.bd
[3] Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
{khandakerasif.ahmed,zakir.hossain}@csiro.au

Abstract. Multiple sclerosis (MS) is a progressive neurodegenerative disease with a wide range of symptoms, making it difficult to diagnose and monitor. Current diagnosis methods are invasive and time-consuming. The use of smartphone monitoring is convenient, non-invasive, and can provide a reliable data source. Our study utilises an open-source dataset, namely—"Floodlight"—that uses smartphones to monitor the daily activities of MS patients. We evaluate whether the Floodlight data can be used in training a machine learning (ML) algorithm for MS diagnosis. After necessary data cleaning, we statistically measured the significance of different tests. Preliminary results show that individual test metrics are helpful for training ML algorithms. Accordingly, we use the selected tests in support vector machine (SVM) and rough set (RS) algorithms. Experimenting with several variations of the ML models, we achieve as high as 69% MS diagnosis accuracy. Since we experiment with SVMs and RSs on individual test metrics, we further report the relative significance of those tests and corresponding ML models suitable for the Floodlight dataset. Our model will serve as a baseline for developing ML-based prognostication tools for MS disease.

Keywords: Multiple sclerosis · Diagnosis · Machine learning · Support vector machine · Floodlight.

1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune and neurodegenerative disease characterised by progressive destruction of the myelin sheath, which insulates nerve cells [2,5]. Damage to nerves of the central nervous system can lead to severe physical and cognitive disability. Symptoms are widely varied and can include blindness, muscle weakness, fatigue, neurological deterioration, and many more [6]. As the disease progresses, these symptoms generally become more severe.

Currently, the precise cause behind MS is still unknown. The wide range of symptoms and unpredictable progression of the disease create further difficulties for diagnosis [4]. Diagnosis of MS is not straightforward and relies on a combination of tests such as blood tests, lumbar puncture, or magnetic resonance imaging (MRI) for lesions [5]. Some of these tests are intrusive but, even then, do not return definitive results.

Although early diagnosis is crucial for treatment, multiple criteria must be met before MS can be diagnosed confidently. For example, one major criterion is that the patient experience two discrete neurological episodes of time and space dissemination [10]. This way of diagnosis is problematic since it requires the patient's MS to be severe enough for this to occur. The expression of this symptom also varies significantly among individuals. These symptoms and episodes may also be unable to be monitored between clinical visits. Another significant criterion is the detection of a lesion via MRI [3]. Again, this requires the disease to have progressed far enough to have a lesion already. Furthermore, detecting lesions is often time-consuming and prone to human error. For these reasons, researchers are using machine learning (ML) algorithms to help with diagnosis [13].

Most current studies that use ML for MS diagnosis utilise MRI scans. In this area, ongoing research is extensive with several available datasets from various clinical settings and prior studies. Numerous methods of data processing and implementations of ML have been explored in this area [18].

Some studies investigate datasets aside from MRIs. Kaur et al. have explored the use of ML to analyse the gaits of MS patients [9]. However, their study is centred around monitoring disease progression, not the diagnosis. It also had a small sample size of 20 MS patients weighted towards older adults, and thus it requires more data to be generalised on different environments. Another study by Pinto et al. used the standard expanded disability status scale (EDSS) and ML to predict the progression of MS [16]. Therefore, existing non-MRI literatures investigated MS prognosis but lacked the diagnosis.

Floodlight Open is a study that uses a smartphone app to collect data from participants over time [14]. The app records information relevant to monitoring MS progression in a noninvasive, consistent, and convenient manner. In this sense, the use of the Floodlight data is unique and advantageous compared to clinical tests since clinical tests may be invasive and cannot monitor patients in-between multiple visits. The app supports several tests that the participants can actively perform daily, such as drawing shapes, pinching, and answering questions. Movements of participants are also passively monitored throughout the day.

Existing studies that use the Floodlight Open data evaluate the extent to which the smartphone app and the Floodlight tests can effectively track MS progression [19]. These studies confirm that smartphone monitoring is an effective way to assess MS progression continuously [12]. Woelfe et al. conducted a study with 264 participants over multiple weeks and found that repetitions of tests can lead to improvements of results due to practice [19]. Our study focuses on using the Floodlight data for diagnosis instead of disease progression monitoring.

We apply ML on the Floodlight data to determine whether these tests are helpful to diagnose MS. Firstly, we analyse the data to understand how it relates to MS diagnosis. Furthermore, we train ML algorithms with the data for predicting MS. This way of predicting MS may add a valuable noninvasive tool for MS diagnosis. An individual should be able to take the Floodlight app tests and input the results into the ML model. The ML model should then return the likelihood of that individual having MS in terms of a percentage score. Among different ML models, support vector machine (SVM) and rough set (RS) showed great potential for different classification tasks.

We utilise SVM and RS algorithms to split data into two groups—individuals with MS or without MS. Data belong to a plot on an n-dimensional space where n is the number of variables related to the groups. The variables are the results from the Floodlight tests as well as background information such as age, sex, and height. A line is drawn between the data points, separating two groups. This line defines which group new data points will fall into, which allows for predictions on whether an individual has MS or not. Compared to other ML algorithms such as random forest or decision tree, SVM is intrinsically suited to two-class problems and draws the line to maximise the margin between the two groups [1]. Hence we chose SVM to diagnose MS from Floodlight data. On the other hand, RS theory uses two sets to approximate the lower and upper bounds of an original set of data [11]. Unlike SVM, which ultimately sorts data into the defined categories (MS or no MS), RS makes it possible to identify borderline cases that do not fit into either lower or upper sets. This potential for "fuzziness" is useful in a medical context because of borderline cases. In a scenario where the SVM algorithm may classify a borderline case as not having MS, the RS algorithm may be able to flag it for further observation.

This paper is structured as follows. Section 2 explains the methodologies with a flow diagram of steps involved in this work. Necessary discussions with the results achieved from data analysis and ML implementation are reported in Section 3. The limitations of this study are explained in Section 4. Finally, we summarised our works with potential future scopes in Section 5.

2 Method

Fig. 1 depicts the step-by-step processes we followed in this study, which are explained in the following subsections.

2.1 Understanding the Data

The Floodlight data is open source and available for download on the website[4]. A unique identifier (ID) is provided for each participant, so it is possible

[4] https://floodlightopen.com/

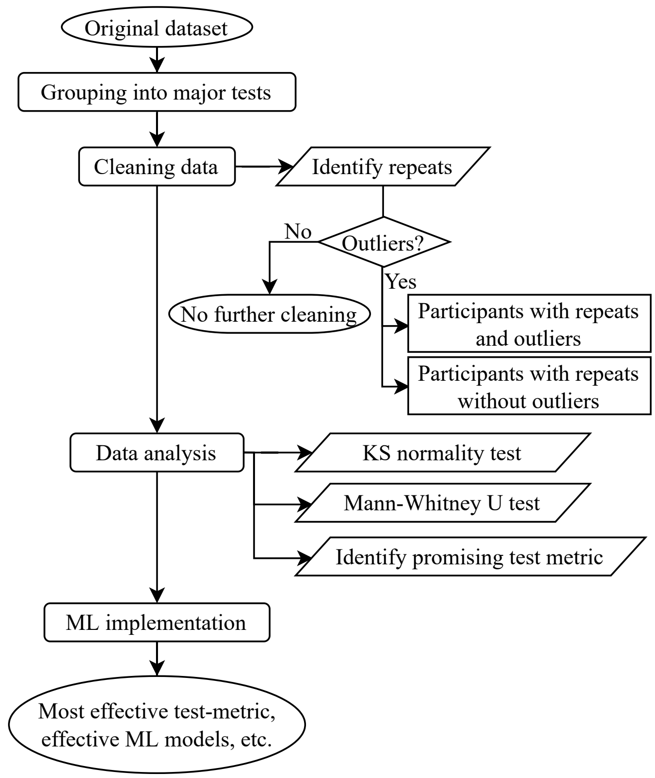

Fig. 1. Steps involved in this work.

to track which tests have been taken and how many times. Background information, including age, weight, sex, height, country of residence, and whether the participant has MS or not, is recorded. While accessing the database, there were 568,600 data points and 2505 unique participants: 1150 participants were without MS, and 1355 participants were diagnosed with MS. The participants (Floodlight app users) self-reported whether they had MS or not. These numbers vary among different test metrics (test metrics refer to different test categories for the users), as detailed in Table 1. Even with this preliminary information, it is possible to identify potential correlations and risk factors associated with certain backgrounds.

Ideal input data for training any ML algorithm is data that very accurately reflects reality. As mentioned previously, MS has varying degrees of severity. To accurately represent this, a continuous spectrum where a higher number indicates more severe MS and a lower number indicates less severe MS would be required. However, the Floodlight data does not reflect this, as participants are only given the option of stating whether they have MS or not. This binary choice leads to some limitations for training the ML algorithm. Due to this lack

Table 1. Number of participants for each test metric.

Test Metric	MS	No MS	Total
Mean Hausdorff Distance	1229	969	2198
Top to Bottom Hausdorff Distance	1223	961	2184
Bottom to Top Hausdorff Distance	1223	960	2183
Circle Hausdorff Distance	1223	960	2183
Square Hausdorff Distance	1223	960	2183
Figure 8 Hausdorff Distance	1223	960	2183
Number of U Turns	1024	774	1798
Average Turn Speed	1024	774	1798
Number of Pinches	1230	981	2211
Daily Mobility Metric	1104	868	1972
1-5 Mood Scale	1297	1077	2374
IPS Correct Responses Baseline	1242	994	2236
IPS Average Response Time Baseline	1242	994	2236

of data, the fuzzy aspect of the RS algorithm was not utilised, and patients were strictly sorted into the MS or no MS to easily compare with SVM performance.

An important distinction needs to be made here. For any given participant, there are two crucial theoretical numbers: how likely the participant has MS versus how severe the participant's MS is. Because the provided data only contains information on whether the patient has MS or not, the ML model will only be able to return the possibility of the participant having MS (and not its severity). However, this possibility will most likely reflect the severity of the participant's MS. For example, if a patient with severe MS takes the Floodlight app tests, this should be reflected in the results. Upon input into the ML model, the model should return that the participant is highly likely to have MS. In contrast, if a patient with mild MS takes the tests, the ML model will probably return that the participant is less likely to have MS compared to the more severe case. In a heuristic sense, the returned likelihood of a patient having MS may reflect the severity of their MS.

2.2 Data Cleaning

The original data source contains a variable called "test name," which describes eight major tests types. These types correspond to actual tests that the participants can perform in the Floodlight app. Each of these tests has one or more corresponding test metrics. This metric is an integer on a continuous scale describing how the participant performed on that test. Each metric has a different range and meaning.

Furthermore, a single test may return multiple test metrics. For example, in the information processing speed (IPS) test, participants are asked to answer a number of questions rapidly. Consequently, two test metrics are returned: the number of correct responses and the average time per response. In the first case, a higher test metric (more correct responses) is a better result, indicating that the participant is less likely to have MS or has less severe MS symptoms. Conversely, for the average response time metric, if the participant has a higher test metric (they take a longer time to respond to each question), the participant is more likely to have MS or has more severe MS symptoms.

Accordingly, we developed a new dataset for each of the eight major test types from the original data for ease of analysis. Within the new dataset, we further split the data into corresponding test metrics described in Table 2.

Table 2. Floodlight test types.

Test Name	**Test Metric**	**Patient's Condition at Higher Score**
Drawing shapes	Number of correct shapes Hausdorff distance difference from correct shape	Better Worse
Pinching test	Number of pinches	Better
U-turn test	Number of U-turns	Better
Daily mood questions	Mood on a scale from 1-5	Better
Sway path and stability	Movement	Worse
Number of steps in 2 minutes	Number of steps taken	Better
Information processing speed (IPS)	Number of correct responses Average time per response	Better Worse
Monitoring of daily mobility[a]	Daily movement	Better

[a]Passively monitored throughout day.

Within each of these test-metric subgroups, null values, zero values, outliers, and other potentially incorrect values were identified and separated. These data points could then be included or excluded. Repeating entries from the same individuals were also handled systematically. If a person repeats a test multiple times, the data for that person should be more accurate; however, it should not lead to the individual being weighted more than others. Thus, repeats from each individual were averaged into a single data point for input into the ML algorithm.

Accordingly, the data within each test-metric was further divided into participants with no repeats and participants with repeats. For participants with only one data point (no repeats), the test metrics' value was assumed to be correct as there is no point of comparison to verify whether the result is abnormal for that participant.

Participants with repeats were separated into two groups. A filter was applied that identified individual participants with potentially incorrect values. Test metrics were plotted for each individual, and it was determined whether they contained any outliers within their own results. The outliers were calculated using the interquartile range (1.5× above or below). The results of the unfiltered participants (no self-contained outliers) were averaged into a single data point.

These three groups—individuals with no repeats, with consistent repeats, and with repeats that have outliers—may be used in combination or individually for inputting into the ML algorithm.

2.3 Preparation of ML Algorithm

For initial screening of which tests have the potential to perform better as input into the ML Algorithm, the Kolmogorov-Smirnov and Mann-Whitney U tests were performed on the cleaned data.

We made two groups for input into the ML algorithm: one group by removing the individuals with outliers, and another by treating them similarly to individuals without outliers (data points of the individual were averaged). Within these two groups, it was also possible to choose whether to use only the test metric (an integer) as the input data or include other variables such as age, height, weight, etc. This gives four groups that can be input into the SVM and RS algorithms: only test metrics without outliers, only test metrics with outliers, all variables without outliers, and all variables with outliers.

The data from the filtered test metrics were then fed into the ML algorithms imported from the python scikit-learn library [15] for SVM and the fuzzy-rough-learn library [11] for RS. There are different options for SVM kernel: linear, sigmoid, Gaussian, or polynomial. Each of the four data groups for each test-metric can be input into these different kernels. Accordingly, we tested 16 SVM and 4 RS variations for each test metric.

We split the data points of each test-metric in an 80/20 ratio: 80% for the training set and 20% for the test set. We used accuracy as our evaluation metric.

3 Results and Discussion

3.1 Exploratory Data Analysis

We analysed the data to identify test metrics that can train the ML algorithm. Ideally, useful test metrics would significantly differ in values when comparing MS patients versus non-MS patients.

Firstly, the data were tested for normality using the Kolmogorov-Smirnov test [7]. We found that most of the Floodlight tests were not normally distributed. For this reason, the nonparametric Mann-Whitney U test [8] was conducted to determine if the differences between MS positive and MS negative test metrics were statistically significant; the results are reported in Table 3 and Fig. 2. For visual ease of comparison, Fig. 2 shows the negative log of the Mann-Whitney U test p values, with the red line showing the $-\log_{10}(0.05)$ threshold.

Table 3. Significance of difference between MS and non-MS patient test metrics.

Test Metric	**Mann-Whitney U Test p Value**
Number of Correct Shapes	8.46×10^{-2}
Mean Hausdorff Distance	6.78×10^{-3}
Top to Bottom Hausdorff Distance	3.44×10^{-2}
Bottom to Top Hausdorff Distance	2.99×10^{-3}
Circle Hausdorff Distance	1.40×10^{-2}
Square Hausdorff Distance	6.53×10^{-4}
Figure 8 Hausdorff Distance	4.85×10^{-5}
Spiral Hausdorff Distance	4.50×10^{-1}
Number of U Turns	7.06×10^{-4}
Average Turn Speed	2.08×10^{-3}
Number of Pinches	3.17×10^{-2}
Daily Mobility Metric	7.95×10^{-10}
Number of Steps Taken[a]	6.38×10^{-6}
Sway Path[a]	9.06×10^{-1}
1-5 Mood Scale	4.06×10^{-29}
IPS Correct Responses Baseline	8.95×10^{-15}
IPS Average Response Time Baseline	4.15×10^{-27}
IPS Correct Responses[a]	3.12×10^{-3}
IPS Average Response Time[a]	3.90×10^{-14}

[a]Test metric result did not reflect expected meaning.

It can be seen that several Floodlight tests have promising p values, indicating that the difference between MS and non-MS patients has the potential to be recognised by an ML algorithm. The most statistically-significant differences are the baseline IPS tests, the mood test, and the daily mobility metric. However, three Floodlight tests—the number of correct shapes, spiral Hausdorff distance, and sway path—showed high p values (> 0.05) and so were not used in the ML implementation.

However, some of these results contradict what the test-metric values are expected to reflect. For example, as shown previously in Table 2, if a patient takes more steps in the 2-minute walk test, this should indicate that they are in a better condition and are less likely to have MS. However, the average number of steps by non-MS patients is actually lower than the non-MS patients in the Floodlight dataset. Therefore, these specific Floodlight tests did not reflect the expected meaning and were thus discarded from inputting into the ML algorithm.

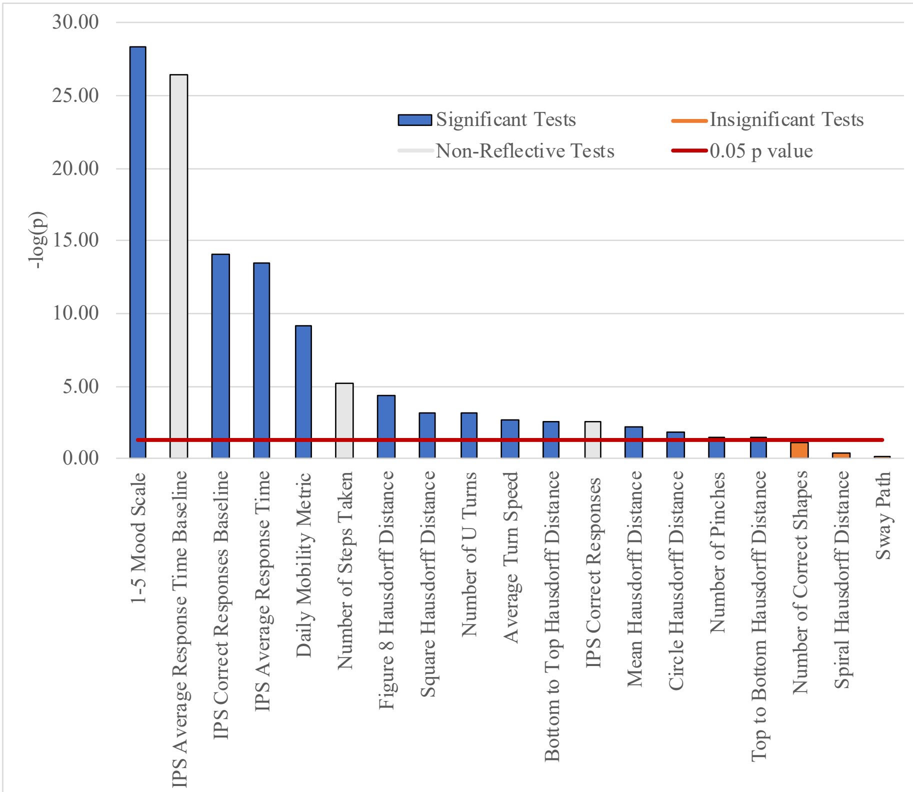

Fig. 2. Tests which have a significant difference between MS and non-MS patients according to the Mann-Whitney U Test.

3.2 ML Results

Table 4 shows the highest accuracy that we achieved for each test metric with corresponding SVM kernel or RS algorithm.

It can be seen that the top to bottom Hausdorff distance had the highest accuracy of 69%, followed by mean Hausdorff distance, square Hausdorff distance, and daily mobility metric, all having an accuracy of 68%.

Regarding the performance of RS compared to SVM, the highest, 2nd, 3rd, 4th, and 5th highest performances were achieved using the RS algorithm. Among SVM kernels, the linear kernel performed the best. In most Hausdorff distance test metrics, the Sigmoid and Gaussian kernels had very similar performances but were poor in comparison to the linear kernel.

Among different data subgroups of whether to consider outliers and all variables, all variables with outliers often gave the best accuracy (Table 4). Therefore, it implies that outliers removal is potentially unnecessary, and patients'

Table 4. Implementation results.

Test Metric	Most Effective Algorithm	Accuracy
Mean Hausdorff Distance[b]	RS	68%
Top to Bottom Hausdorff Distance[b]	**RS**	**69%**
Bottom to Top Hausdorff Distance[b]	RS	67%
Circle Hausdorff Distance[a]	RS	67%
Square Hausdorff Distance[a]	RS	68%
Figure 8 Hausdorff Distance[a]	RS	65%
Number of U Turns[b]	Linear SVM	63%
Average Turn Speed[b]	Linear SVM	67%
Number of Pinches[b]	Linear SVM	64%
Daily Mobility Metric[a]	RS	68%
1-5 Mood Scale[b]	Linear SVM	64%
IPS Correct Responses Baseline[a]	RS	64%
IPS Average Response Time Baseline[b]	Linear	65%

[a]All variables without outliers.
[b]All variables with outliers.
[c]Only test metric without outliers.
[d]Only test metric with outliers.
Acronyms: RS – rough set, SVM – support vector machine.

background information, such as age, height, weight, etc., positively influences ML-based MS diagnosis.

We could not directly compare the outcomes of this study to related literature since, to the best of our knowledge, there is no existing research that has similarities to the test metrics used.

4 Limitation

As mentioned previously, the severity of a patient's MS is not recorded in the provided Floodlight dataset. This led to the ML algorithm only being able to return true or false as a prediction for whether the patient had MS. This scenario means that a patient with very severe MS and another with very mild MS would both output true with no way to distinguish between the two in terms of certainty. If it were possible to obtain data on the severity of each participant's MS, for example, through an EDSS score, it might be possible to distinguish between these two cases, especially using the RS algorithm.

It is worthwhile to note that the current accuracies are achieved using only single test metric within the dataset. For example, the 69% accurate prediction is achieved using data only from the top to bottom Hausdorff distance and nothing else. Thus, combining all Floodlight tests will most likely increase prediction accuracy.

Even though MRI-based studies to detect neural damage or demyelinating changes in the brain are proven as gold-standard to detect MS, we aimed to develop a cost-efficient ML model based on physiological signals, which will pave the way for future MS prognostication. Physical and physiological signals have shown great potential for monitoring mental health [17], and our study also prove their potential for MS detection.

5 Conclusion and Future Work

This study has investigated the potential of utilising the tests from the Floodlight smartphone app for diagnosing MS. We found that multiple test metrics could provide information that may distinguish between MS and non-MS patients. First, the Floodlight data were cleaned from repetitions and outliers. After applying two statistical tests—Kolmogorov-Smirnov and Mann-Whitney U test—we acquired statistically significant data that we used in SVM and RS algorithms. Experimenting with four different SVM kernels and RS algorithm on each of the selected tests individually, we achieved a maximum of 69% accuracy by a RS model in diagnosing MS patients using only a single test metric. We report top to bottom Hausdorff distance as the most effective Floodlight test metric. This research throw light on using smartphone monitoring data in ML-based MS diagnosis. Future research might focus on further refinements, such as implementing other ML algorithms like multilayer perceptron or decision trees, with further tuning on hyperparameters using algorithms like grid search method. Also, given a data set with information on the degree of severity of MS, it may be possible to extend the application of RS "fuzziness." Moreover, since this study has focused on investigating each test-metric separately, we may look into using multiple combinations or all of them in conjunction.

References

1. Abdullah, D.M., Abdulazeez, A.M.: Machine learning applications based on SVM classification a review. Qubahan Academic Journal **1**(2), 81–90 (Apr 2021). https://doi.org/10.48161/qaj.v1n2a50
2. Chaudhuri, A.: Multiple sclerosis is primarily a neurodegenerative disease. Journal of neural transmission **120**(10), 1463–1466 (2013). https://doi.org/10.1007/s00702-013-1080-3
3. Eshaghi, A., Young, A.L., Wijeratne, P.A., Prados, F., Arnold, D.L., Narayanan, S., Guttmann, C.R.G., Barkhof, F., Alexander, D.C., Thompson, A.J., Chard, D., Ciccarelli, O.: Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. Nature Communications **12**(2078), 1–12 (Apr 2021). https://doi.org/10.1038/s41467-021-22265-2
4. Gelfand, J.M.: Chapter 12 - multiple sclerosis: diagnosis, differential diagnosis, and clinical presentation. In: Goodin, D.S. (ed.) Multiple Sclerosis and Related Disorders, Handbook of Clinical Neurology, vol. 122, pp. 269–290. Elsevier (2014). https://doi.org/10.1016/B978-0-444-52001-2.00011-X

5. Ghasemi, N., Razavi, S., Nikzad, E.: Multiple sclerosis: pathogenesis, symptoms, diagnoses and cell-based therapy. Cell Journal (Yakhteh) **19**(1), 1–10 (2017). https://doi.org/10.22074/cellj.2016.4867
6. Goldenberg, M.M.: Multiple sclerosis review. P & T **37**(3), 175–184 (Mar 2012)
7. Gupta, A., Mishra, P., Pandey, C., Singh, U., Sahu, C., Keshri, A.: Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia **22**(1), 67–72 (2019). https://doi.org/10.4103/aca.aca_157_18
8. Hart, A.: Mann-whitney test is not just a test of medians: differences in spread can be important. BMJ **323**(7309), 391–393 (Aug 2001). https://doi.org/10.1136/bmj.323.7309.391
9. Kaur, R., Chen, Z., Motl, R., Hernandez, M.E., Sowers, R.: Predicting multiple sclerosis from gait dynamics using an instrumented treadmill: A machine learning approach. IEEE Transactions on Biomedical Engineering **68**(9), 2666–2677 (Sep 2021). https://doi.org/10.1109/tbme.2020.3048142
10. Leary, S.M.: Multiple sclerosis: diagnosis and the management of acute relapses. Postgraduate Medical Journal **81**(955), 302–308 (May 2005). https://doi.org/10.1136/pgmj.2004.029413
11. Lenz, O.U., Peralta, D., Cornelis, C.: fuzzy-rough-learn 0.1: A python library for machine learning with fuzzy rough sets. In: Bello, R., Miao, D., Falcon, R., Nakata, M., Rosete, A., Ciucci, D. (eds.) Rough Sets. pp. 491–499. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-52705-1_36
12. Midaglia, L., Mulero, P., Montalban, X., Graves, J., Hauser, S.L., Julian, L., Baker, M., Schadrack, J., Gossens, C., Scotland, A., Lipsmeier, F., van Beek, J., Bernasconi, C., Belachew, S., Lindemann, M.: Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: Nonrandomized interventional feasibility study. Journal of Medical Internet Research **21**(8), e14863 (Aug 2019). https://doi.org/10.2196/14863
13. Moazami, F., Lefevre-Utile, A., Papaloukas, C., Soumelis, V.: Machine learning approaches in study of multiple sclerosis disease through magnetic resonance images. Frontiers in Immunology **12** (Aug 2021). https://doi.org/10.3389/fimmu.2021.700582
14. Montalban, X., Graves, J., Midaglia, L., Mulero, P., Julian, L., Baker, M., Schadrack, J., Gossens, C., Ganzetti, M., Scotland, A., Lipsmeier, F., van Beek, J., Bernasconi, C., Belachew, S., Lindemann, M., Hauser, S.L.: A smartphone sensor-based digital outcome assessment of multiple sclerosis. Multiple Sclerosis Journal **28**(4), 654–664 (2022). https://doi.org/10.1177/13524585211028561
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
16. Pinto, M.F., Oliveira, H., Batista, S., Cruz, L., Pinto, M., Correia, I., Martins, P., Teixeira, C.: Prediction of disease progression and outcomes in multiple sclerosis with machine learning. Scientific Reports **10**(21038) (Dec 2020). https://doi.org/10.1038/s41598-020-78212-6
17. Rostov, M., Hossain, M.Z., Rahman, J.S.: Robotic emotion monitoring for mental health applications: Preliminary outcomes of a survey. In: Human-Computer Interaction – INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30 – September 3, 2021, Proceedings, Part V. p. 481–485. Springer-Verlag, Berlin, Heidelberg (2021). https://doi.org/10.1007/978-3-030-85607-6_62

18. Shoeibi, A., Khodatars, M., Jafari, M., Moridian, P., Rezaei, M., Alizadehsani, R., Khozeimeh, F., Gorriz, J.M., Heras, J., Panahiazar, M., Nahavandi, S., Acharya, U.R.: Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review. Computers in Biology and Medicine **136**, 104697 (Sep 2021). https://doi.org/10.1016/j.compbiomed.2021.104697
19. Woelfle, T., Pless, S., Wiencierz, A., Kappos, L., Naegelin, Y., Lorscheider, J.: Practice effects of mobile tests of cognition, dexterity, and mobility on patients with multiple sclerosis: Data analysis of a smartphone-based observational study. J Med Internet Res **23**(11), e30394 (Nov 2021). https://doi.org/10.2196/30394