Detection of Bruxism Using Data From an In-Mouth Accelerometer

Using Hidden Markov Models to detect bruxism events in intraoral data

Floris van der Voorn[1]

Supervisor(s): Przemysław Pawełczak[1]**, Vivian Dsouza**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Floris van der Voorn
Final project course: CSE3000 Research Project
Thesis committee: Przemysław Pawełczak, Vivian Dsouza, Johan Pouwelse

Abstract

Bruxism is a medical disorder that causes individuals to grind or clench their teeth together. This can cause dental damage, headaches, and jaw disorders. Reliable detection of bruxism remains challenging due to the limitations of current diagnostic methods. This study investigates the potential of using an intraoral device equipped with an accelerometer to detect bruxism events. Machine learning techniques, specifically Hidden Markov Models, were applied to classify accelerometer data collected during controlled experiments. The results indicate that while the system effectively distinguishes between bruxism and non-bruxism activities in general, detecting actual grinding and clenching events remains difficult due to class imbalance and overlapping motion patterns. Nonetheless, the findings suggest that in-mouth accelerometers hold promise for future bruxism detection systems, requiring further data collection and model refinement.

1 Introduction

The mouth is a valuable resource for medical data from people. It serves as a method for monitoring overall health, with conditions such as sleep apnea, bruxism (teeth grinding), dehydration, and infections often manifesting early signs through oral indicators. Recognizing the potential of oral data, researchers created the Densor.[1] The Densor is a compact device designed to record data from the mouth in real time. It collects data using different sensors from barometer to accelerometer data.

To harness the power of the Densor this research will try to see if it is possible to detect bruxism using the accelerometer data form the Densor. Bruxism is a medical term for clenching and grinding of the teeth. It is one of the more common involuntary habits, occurring in 8-12% of the general population.[2] It can occur both while awake and asleep. It can cause multiple adverse health problems such as chipped teeth, headaches, poor sleep, and jaw pain. The main cause of bruxism is not yet known, but there appears to be a genetic component to it.[3] It has also been linked to stress or anxiety of a person.

There are multiple ways to diagnose bruxism. For example: polysomnography (PSG) or electromyogra- phy (EMG). But they are either unreliable or difficult to perform. Therefore, finding a good way to detect bruxism is required. This research uses data collected by a Densor, to find if it is possible to be used to diagnose bruxism. This research will use Hidden Markov Models (HMMs) to detect both grinding and clenching from the collected data.

The main question of this research is: Can a Densor be used to detect bruxism events? This will be split up in two questions:

- Can grinding be detected from the accelerometer data with an 80% accuracy?
- Can clenching be detected from the accelerometer data an 80% accuracy?

2 Related Works

There are different methods used for diagnosing bruxism. These include patient self-reports, the use of EMG and using accelerometers. The gold standard however is a PSG analysis.[4]These methods however still suffer from limitations such as inaccuracy, user comfort and cost.

The PSG is the best because it can be used to make a quantitative assessment of the tongue, mouth and jaw movements. It uses multi-modal measurements of such parameters as electroencephalography (EEG), EMG, electrocardiography (ECG), air flow monitoring and audio–video recording. These can then be used to make an accurate diagnosis. There are however still problems. A PSG is a high cost methods that takes multiple nights to preform. And although they can be done at home they are mostly done in sleep labs. This may not be representative of a natural sleep milieu, which may hamper diagnosis.

The self reporting of patients is a common method used to diagnose bruxism early on in the process. This method requires patients or their family or bed partners to witness the bruxism. The symptoms of the patient can also be used in the diagnosis. These methods however are unreliable, having a low to medium correlation with instrumental approaches.[5] They can also be required to fill in questionnaires about their symptoms.

Portable EMG devices can also be used for diagnosing bruxism.[6] They are used for detecting muscle movements of the masseter muscle. These are mounted on the outside of the face on the chin and jaws. They measure the amount of bursts of the EMG. For accurate results they require to be worn 3 to 5 nights. They also require patients to do a setup procedure by relaxing and clenching their jaw for a sort period of time. The accuracy of the diagnosis comes close to the diagnosis of PSG analysis.

Accelerometers, commonly used in sleep and physical activity tracking, have also been used to monitor jaw and head movements associated with bruxism. There have been some studies that showed that using a accelerometer on the side or in the ear of patients might be able to detect bruxism. A study showed that wearing in ear accelerometer and gyroscope could be used.[7]

Another study also showed using a prototype that bruxism events might be detectable by accelerometers on the face. The use of in-mouth accelerometers for bruxism has not been well explored. This approach benefits from proximity to the source of bruxism activity, enabling high-resolution motion capture with minimal external interference. However, challenges remain in distinguishing bruxism from other oral movements such as speaking or swallowing.

3 Methodology

In this part the data collection will first be discussed. Then how the data is labeled will be shown. Then the feature extraction will be explained. Lastly the classification will be done.

3.1 Densor Data Collection

The Densor contains multiple sensors that could be used to gather information. But for this research only the inertial Measurement Unit (IMU) was used. Since the Densor is an intraoral device, it is designed to be small for the comfort and safety of the user. This does have its limitations: The lifespan of the Densor is limited. The team has done significant work to make it as efficient as possible, but the data gathered during a full night still has a low frequency. This can be detrimental for detecting bruxism. Thus, the data used for the research was done in small time windows and has a frequency of 200 Hz. This has been done to look into the feasibility to detect bruxism using an intraoral accelerometer. Data for this project was provided by the Densor team. The data contained 3 columns for the x y and z axes of the acceleration. The data was gathered by being in different positions and doing tasks when prompted. An experiment looked like this:

- Wait some time until prompted
- Do the action until the prompt disappears.
- Wait until prompted again.
- Do the action until the prompt disappears.
- Wait until 20 seconds are over.

This setup automatically labeled the raw data as either "performing the action" or "not performing the action." Each type of activity was recorded multiple times, resulting in 300 data segments of 20 seconds each. Among these, 8 segments contained grinding activity and 8 segments contained clenching activity.

An example of the recorded data and its corresponding labels can be seen in Figure 1, which shows the acceleration along the x, y, and z axes during an experimental trial. The highlighted regions represent the periods when the subject was actively performing the prompted action.

3.2 Labeling

To use this data for machine learning algorithms it was made into shorter windows. The raw data was segmented into sliding windows of 1 second with a 0.5 second overlap. These were then classified based on the majority of the classification of the points in the data. This was done to correctly catch the transition from grinding or clenching into not grinding or clenching. This data was then adapted into features and used to train machine learning algorithms.

3.3 Feature extraction

For each axis separately, a predefined set of statistical and signal features was calculated. These included the mean, standard deviation, minimum, and maximum values to describe the central tendency and variability of the signal. The number of zero-crossings was computed to quantify the rate of signal sign changes, which is indicative of oscillatory behavior. Root mean square (RMS) values were used to represent the signal's energy,. Additionally, peak-based features such as the total number of peaks and the average peak height were extracted.

Features were also extracted from the signal magnitude, calculated as the Euclidean norm of the three axes. These

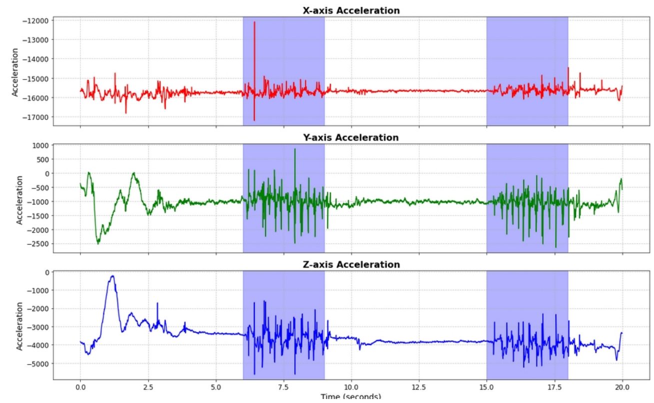

Figure 1: Raw accelerometer data from an experiment involving teeth grinding, with blue regions indicating periods of active grinding.

magnitude-based features included the signal magnitude area (SMA) and the variance of the magnitude itself. Further features included the number of peaks and the mean peak height of the magnitude signal.

These features where extracted using the scipy and numpy libraries in python.

3.4 Classification

Two machine learning algorithms were used to classify and get the results for the project: random forest and HMMs. Common python libraries scikitlearn and hmm were used for the machine learning algorithms. To train the machine learning algorithms the data was first split using a 5 fold cross validation. To compare these algorithms the accuracy, precision, recall, and f1-score were used. In this paper only the results of the HMMs will be shown as it had the better results.

4 Experimental Setup

This section describes the experimental procedure used to get the results shown in this paper. This procedure is repeated twice. Once for grinding classification and once for clenching classification.

4.1 Activity Classification

All collected sensor data described in Section 3 was utilized for the experiment. The following steps summarize the procedure:

1. **Data Preprocessing and Windowing**
   The multivariate sensor data, consisting of the x-axis, y-axis, z-axis, and calculated magnitude, was segmented into overlapping windows. Each window had a duration of 1 second with a stride of 0.5 seconds.
2. **Feature Extraction**
   Within each window, statistical features were extracted separately for each of the four channels. The features were selectively enabled based on the correlation between the activity and the feature.
3. **Cross-Validation**
   A stratified 5-fold cross-validation scheme was applied

to ensure balanced representation of both classes in each fold. Then two models where trained for both classes.

4. **Prediction and Evaluation**
   Each sample in the test set was evaluated by computing its log-likelihood under both HMMs. The final predicted label corresponded to the model with the higher likelihood. All predictions across the folds were collected for overall evaluation.

5. **Global Evaluation**
   After cross-validation, the following performance metrics were calculated across the entire dataset:

   - Precision
   - Recall
   - F1-score

   Additionally, a confusion matrix was generated to visualize the overall classification performance.

5 Results

First the grinding results will be viewed. Then the results of the clenching classifier will be shown. The code for this can be found in the repository[8].

5.1 Grinding

The classifier for grinding achieved an overall accuracy of 98%, with a weighted F1-score of 0.98 across both classes. This, however, does not tell the whole story. There was a significant class imbalance present. The nothing class constituted the vast majority of the data (7568 samples), while the grinding class was much smaller (93 samples).

For the nothing class, the precision, recall, and F1-score were extremely high (1.00, 0.98, and 0.99 respectively). However, for the grinding class, performance was way lower with a precision of 0.30, recall of 0.67, and F1-score of 0.41. This suggests that while the model correctly identified the majority of nothing instances, its ability to detect grinding was limited.

The confusion matrix (Figure 2) illustrates this imbalance, showing 62 true positives for grinding, but also 31 false negatives, and 148 false positives, where nothing instances were incorrectly classified as grinding. The detailed classification report is shown in Figure 3.

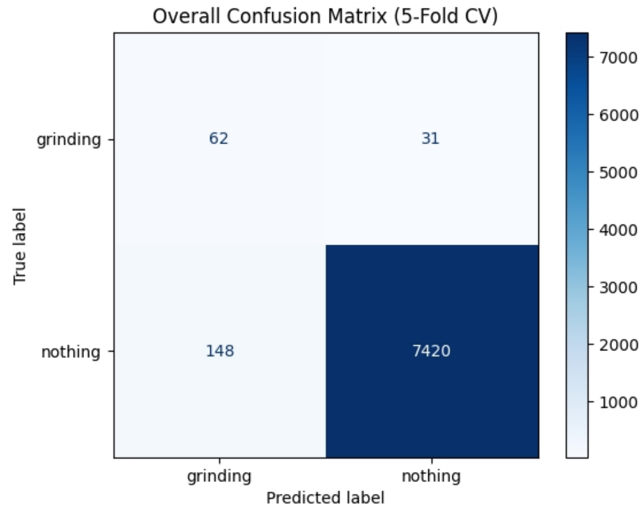

Figure 2: Confusion matrix for grinding classification.

```
=== Overall Classification Report (5-Fold CV) ===
              precision    recall  f1-score   support

     nothing       1.00      0.98      0.99      7568
    grinding       0.30      0.67      0.41        93

    accuracy                           0.98      7661
   macro avg       0.65      0.82      0.70      7661
weighted avg       0.99      0.98      0.98      7661
```

Figure 3: Classification report table for grinding classification.

5.2 Clenching

Similarly to the grinding classification, the classifier achieved an overall accuracy of 96%, with a weighted F1-score of 0.97 across both classes. However, this high-level performance is again influenced by a significant class imbalance. The nothing class dominated the dataset with 6771 samples, while the clenching class was heavily underrepresented with only 93 samples.

For the nothing class, the classifier demonstrated excellent performance, achieving a precision of 1.00, recall of 0.96, and F1-score of 0.98. In contrast, the clenching class showed a lower performance, with a precision of 0.21, recall of 0.80, and F1-score of 0.33. This means that the model was relatively good at identifying most clenching events, but also tended to misclassify a high amount of nothing instances as clenching, leading to many false positives.

The confusion matrix (Figure 4) illustrates this disparity: of the 93 true clenching instances, 74 were correctly predicted, while 19 were misclassified as nothing. Conversely, among the 6771 nothing instances, 283 were incorrectly labeled as clenching, with 6488 correctly identified. The corresponding classification report table is shown in Figure 5.

6 Responsible Research

This research involved the use of sensor data collected from human subjects performing specific mouth-related activities. All data collection was conducted under controlled laboratory conditions with informed consent from the participants involved. The collected data did not contain any personally identifiable information, ensuring the privacy and anonymity of the individuals.

One of the ethical considerations in this research is about the future use of intraoral devices for bruxism detection in real-life. While the proposed system may support diagnosing bruxism, it is not a replacement for clinical assessment by qualified professionals. People should not be relying only on the automated systems and should look for professional help from appropriate sources.

Regarding reproducibility, the methodology used in this study was fully based on standard, publicly available machine

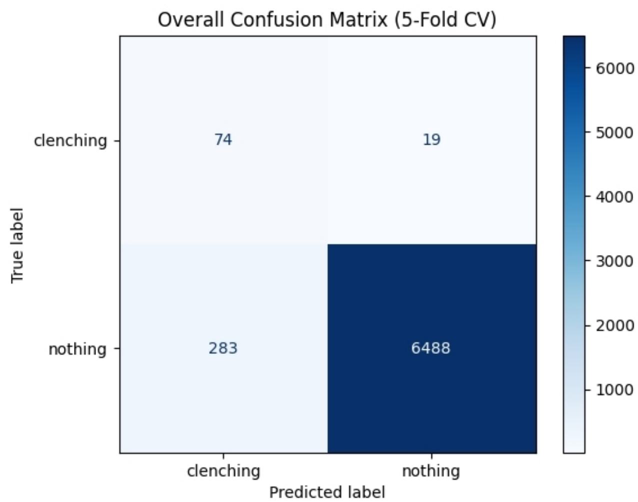

Figure 4: Confusion matrix for clenching classification.

=== Overall Classification Report (5-Fold CV) ===

	precision	recall	f1-score	support
nothing	1.00	0.96	0.98	6771
clenching	0.21	0.80	0.33	93
accuracy			0.96	6864
macro avg	0.60	0.88	0.65	6864
weighted avg	0.99	0.96	0.97	6864

Figure 5: Classification report table for clenching classification.

learning libraries and signal processing techniques. Data preprocessing steps, feature extraction procedures, and classification algorithms are described in detail to enable replication. The feature extraction process, statistical feature calculation, and peak analysis, was implemented using standard Python packages such as NumPy and SciPy. The HMM classifier was trained and evaluated using commonly used software libraries.

However, reproducibility may be affected by the limited availability of the original dataset. The dataset was collected using a specific prototype of the Densor device under controlled conditions, and access to this data is restricted. As such, researchers seeking to replicate or extend this work would need to conduct their own data collection under similar experimental conditions. Future work should consider the creation of an open dataset to support transparent and reproducible research in this domain.

7 Discussion

The results show that bruxism-related activities, including both grinding and clenching, can potentially be detected using in-mouth accelerometer data processed by Hidden Markov Models. The models can for the most part detect when no grinding or clenching is going on. They have a more difficult time with distinguishing when the action is actually occurring.

For both grinding and clenching, the classifiers showed excellent performance on the dominant class, with high precision, recall, and F1-scores. Whereas, detection of the actual bruxism activities was less reliable. The grinding class suffered from low precision (0.30) and moderate recall (0.67), while the clenching class showed similarly poor precision (0.21) despite a higher recall (0.80). These low precision scores indicate a substantial number of false positives in both cases. A likely cause for this is the class imbalance in the dataset.

A significant challenge of this study was indeed the limited amount of labeled bruxism activity data. This imbalance likely biased the classifier towards predicting the majority class, leading to reduced precision for both grinding and clenching detection. Furthermore, the similarity of accelerometer patterns between these activities and other mouth movements may have contributed to the model's difficulty in discriminating the classes without incorporating additional features.

It is also important to note that data collection occurred in a controlled, short-duration environment, which may not fully represent the complexity and variability of real-world bruxism behaviors, especially those occurring unconsciously during sleep.

8 Conclusions

This study explored the use of an intraoral device, the Densor, equipped with an accelerometer to detect bruxism-related activities such as teeth grinding and clenching. Using data collected from controlled laboratory experiments, features were extracted from the accelerometer signals, and Hidden Markov Models were trained to classify these activities.

The results demonstrated high overall accuracy and strong detection of non-bruxism activity, reflecting the model's effectiveness in identifying when no grinding or clenching occurred. However, detection of actual bruxism events was less reliable. The classifiers exhibited low precision and produced a substantial number of false positives when attempting to identify grinding and clenching actions. This performance issue was largely attributed to a severe imbalance in the dataset, where non-bruxism data significantly outweighed bruxism event samples. Furthermore, the similarity of oral motion patterns, such as speaking or swallowing, complicated accurate classification.

Ultimately, the study indicates that while in-mouth accelerometers show potential for detecting bruxism, the method in its current form faces challenges in reliably distinguishing bruxism events from other oral movements, particularly under controlled experimental conditions with limited data diversity.

9 Future Work

While this study showed that using intraoral accelerometers to detect bruxism has potential, there are still several important areas for future research to make these systems more reliable, accurate, and practical.

One major issue is the imbalance in the data collected. There were far fewer examples of grinding and clenching compared to the nothing class. This made it harder for the classifier to learn to detect actual bruxism events. Future research should focus on gathering a much larger and more balanced set of data. This should include data from people who have been diagnosed with bruxism. Collecting data during long periods, especially overnight while people are asleep, would help capture natural grinding and clenching behavior. Real nighttime data from actual bruxism patients would give more useful and realistic examples for training the detection system. This would help the machine learning model better recognize real-life bruxism activity.

In terms of machine learning methods, while HMMs were helpful in this study, they may not be the best choice for capturing the complex patterns in this kind of data. Future work could explore more advanced machine learning techniques, such as deep learning models like convolutional or recurrent neural networks. These are designed to handle time-based and sequence data very well. They could also learn directly from the raw signals of the sensors, reducing the need for manual feature selection and making the system more adaptable to differences in how people behave.

Finally, future research should test the system in real-life situations and not just in short, controlled lab sessions. For example, it should be tested in people's homes, over several nights, to see how well it works in a normal environment. This would show how the system performs when faced with everyday challenges like different sleeping habits, changes in device position, and other interruptions. Longer-term studies could also check if the system stays reliable over time, even as the user gets tired or the device shows wear and tear.

References

[1] V. Dsouza, J. Pronk, C. Peppelman, V. I. Madariaga, T. Pereira-Cenci, B. Loomans, and P. Pawełczak, "Densor: An intraoral battery-free sensing platform," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 4, nov 2024. [Online]. Available: https://doi.org/10.1145/3699746

[2] H. Kanathila, A. Pangi, B. Poojary, and M. Doddamani, "Diagnosis and treatment of bruxism: Concepts from past to present," *Int J Appl Dent Sci*, vol. 4, no. 1, pp. 290–295, 2018.

[3] H. Beddis, M. Pemberton, and S. Davies, "Sleep bruxism: an overview for clinicians," *British Dental Journal*, vol. 225, no. 6, pp. 497–501, 2018. [Online]. Available: https://doi.org/10.1038/sj.bdj.2018.757

[4] G. J. LAVIGNE, S. KHOURY, S. ABE, T. YAMAGUCHI, and K. RAPHAEL, "Bruxism physiology and pathology: an overview for clinicians," *Journal of Oral Rehabilitation*, vol. 35, no. 7, pp. 476–494, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2842.2008.01881.x

[5] W. Yachida, T. Arima, E. E. Castrillon, L. Baad-Hansen, N. Ohata, and P. Svensson, "Diagnostic validity of self-reported measures of sleep bruxism using an ambulatory single-channel emg device," *Journal of Prosthodontic Research*, vol. 60, no. 4, pp. 250–257, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1883195816000074

[6] M. Maeda, T. Yamaguchi, S. Mikami, W. Yachida, T. Saito, T. Sakuma, H. Nakamura, M. Saito, M. Mizuno, K. Yamada, and K. Satoh, "Validity of single-channel masseteric electromyography by using an ultraminiature wearable electromyographic device for diagnosis of sleep bruxism," *Journal of Prosthodontic Research*, vol. 64, no. 1, pp. 90–97, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1883195818304626

[7] E. Bondareva, E. R. Hauksdottir, and C. Mascolo, "Earables for detection of bruxism: A feasibility study," in *UbiComp/ISWC 2021 - Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 146–151. [Online]. Available: https://doi.org/10.1145/3460418.3479327

[8] F. van der Voorn, "Bruxism densor," https://github.com/florisv57/Bruxism_densor/, 2025, accessed: 2025-06-22.