

Original Paper

# Prediction of Mini-Mental State Examination Scores for Cognitive Impairment and Machine Learning Analysis of Oral Health and Demographic Data Among Individuals Older Than 60 Years: Cross-Sectional Study

Alper Idrisoglu, MSc; Johan Flyborg, PhD, DDS; Sarah Nauman Ghazi, MSc; Elina Mikaelsson Midlöv, MSc; Helén Dellkvist, MSc; Anna Axén, MSc; Ana Luiza Dallora, PhD

Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden

**Corresponding Author:**

Alper Idrisoglu, MSc  
Department of Health  
Blekinge Institute of Technology  
Valhallavägen 1  
Karlskrona 371 41  
Sweden  
Phone: 46 701462619  
Email: [alper.idrisoglu@bth.se](mailto:alper.idrisoglu@bth.se)

## Abstract

**Background:** As the older population grows, so does the prevalence of cognitive impairment, emphasizing the importance of early diagnosis. The Mini-Mental State Examination (MMSE) is vital in identifying cognitive impairment. It is known that degraded oral health correlates with MMSE scores  $\leq 26$ .

**Objective:** This study aims to explore the potential of using machine learning (ML) technologies using oral health and demographic examination data to predict the probability of having MMSE scores of 30 or  $\leq 26$  in Swedish individuals older than 60 years.

**Methods:** The study had a cross-sectional design. Baseline data from 2 longitudinal oral health and ongoing general health studies involving individuals older than 60 years were entered into ML models, including random forest, support vector machine, and CatBoost (CB) to classify MMSE scores as either 30 or  $\leq 26$ , distinguishing between MMSE of 30 and MMSE  $\leq 26$  groups. Nested cross-validation (nCV) was used to mitigate overfitting. The best performance-giving model was further investigated for feature importance using Shapley additive explanation summary plots to easily visualize the contribution of each feature to the prediction output. The sample consisted of 693 individuals (350 females and 343 males).

**Results:** All CB, random forest, and support vector machine models achieved high classification accuracies. However, CB exhibited superior performance with an average accuracy of 80.6% on the model using  $3 \times 3$  nCV and surpassed the performance of other models. The Shapley additive explanation summary plot illustrates the impact of factors on the model's predictions, such as age, Plaque Index, probing pocket depth, a feeling of dry mouth, level of education, and use of dental hygiene tools for approximal cleaning.

**Conclusions:** The oral health parameters and demographic data used as inputs for ML classifiers contain sufficient information to differentiate between MMSE scores  $\leq 26$  and 30. This study suggests oral health parameters and ML techniques could offer a potential tool for screening MMSE scores for individuals aged 60 years and older.

*JMIR Med Inform* 2025;13:e75069; doi: [10.2196/75069](https://doi.org/10.2196/75069)

**Keywords:** classification; machine learning; mini-mental state examination; cognitive impairment; oral health

## Introduction

The anticipated global population of 8 billion individuals for 2023 has been achieved. The population growth continues to persist, albeit experiencing a decelerated trajectory. Global birth rates are declining, juxtaposed with a consistent upward trajectory in average life expectancy, attributed to advancements in scientific knowledge, improved nutritional standards, advancements in public health infrastructure, and enhanced sanitation practices [1]. An estimate for the year 2050 suggests that individuals aged 60 years and older will constitute 38% of the total population and will continue to increase. While many individuals enjoy several quality years of life, a significant portion of them will also struggle with neurocognitive diseases such as Alzheimer disease (AD) and mild cognitive impairment (MCI) [2]. Until now, diagnosing AD, which is the most common neurocognitive disorder, has required extensive hospital-level efforts, and no curative treatment has been available [3]. Recent research has identified biomarkers in blood that identify proteins associated with AD [4]. At the same time, several approved antibody medications now stop the breakdown of nerve cells caused by phosphorylated Tau and amyloid plaques [5]. This is expected to significantly burden the health care system, as early diagnosis is essential for a good treatment result. The need for screening methods is expected to be extensive. It is widely recognized that MCI and neurocognitive diseases contribute to deteriorating oral health, with individuals affected by these conditions exhibiting poorer oral health [6,7]. The decline in oral health carries profound implications, causing hardship for affected individuals and placing substantial demands on society and health care systems in terms of both costs and human resources. However, recent research suggests that the deterioration of oral health in individuals with MCI can be delayed by at least 12 months through the introduction of a powered toothbrush [8]. Additionally, severe periodontitis poses a risk factor for dementia [9]. Various instruments are available to assess cognitive function, with one of the most commonly used being the Mini-Mental State Examination (MMSE) test, which is internationally recognized and validated in numerous languages. The MMSE scale ranges from 0 to 30, with higher scores indicating a better cognitive function [10]. Recently, machine learning (ML) has emerged as a valuable aid in enhancing the precision and efficiency of cognitive assessments [11,12], leading to significant progress in the field of cognitive research [13-15]. Various datasets and algorithms can be used to better predict MMSE scores, differentiate between MCI and AD, and improve cognitive test batteries. A comprehensive study used a multifaceted approach, integrating brain cortical characteristics, biological markers, risk factors, positron emission tomography scan measures, and cognitive scores into a joint feature set to predict MMSE scores at 6 and 12 months, achieving mean absolute errors of 1.40 [16]. Researchers have improved predictive accuracy by experimenting with ML algorithms, including logistic regression, decision trees, support vector machine (SVM), XGBoost, and random forests (RF). This approach enhances diagnostic precision and offers valuable

insights into disease progression and treatment response [17]. Similarly, previous research has explored integrating MRI data with cognitive tests, such as logical memory, to detect and distinguish individuals with normal cognition from those with MCI. Through deep learning models trained on MRI slices and fusion techniques, the combined model surpassed individual modalities in predictive performance, underscoring the potential of multimodal fusion in cognitive assessment, achieving an accuracy of 90.9% [18]. Earlier research has explored the area of oral health and its relationship with cognitive health. One meta-analysis [19] found that oral health in individuals with dementia was significantly worse than that of the controls. Another systematic literature review [20] came to a similar conclusion, attributing this to both difficulties in self-care due to the consequences of the disease and inflammatory mechanisms. Further, a systematic literature review [21] also indicated an association between oral health and cognitive impairment, implying a bidirectional relationship between these and ratifying a need for higher-level evidence in this area. Another systematic literature review analyzed the effects of oral interventions on cognition and found that dental treatments had a subjective influence on cognition [22]. One study [23] explored the relationship between oral health-related quality of life and cognitive function among individuals in residential care settings. Significant correlations were found between poor oral health and cognitive impairment, including MCI, revealing a relationship between cognitive status and perceived oral health [23]. Another study investigated the use of noninvasive digital biomarkers to quantify oral health and applied ML algorithms to detect cognitive decline within a community setting. Significant findings included notable differences in oral diadochokinesis rates and oral acidity between individuals with cognitive decline and those with normal cognitive function. This further demonstrates that digital oral health biomarkers have the potential to track cognitive function and facilitate early detection of cognitive decline [24].

Despite promising advancements in clinical practice and the possibility of early diagnosis, further research is needed. One significant research gap is the lack of exploration of oral health parameters as potential indicators for estimating MMSE scores. Although oral health has a recognized influence on overall well-being and cognitive function, there have been no empirical studies to explore the integration of oral health metrics into ML-based cognitive assessment models. This opens an opportunity for future studies to expand the scope of cognitive assessments and develop more comprehensive diagnostic frameworks. Many individuals routinely visit dental care facilities for treatment and preventive care. Leveraging existing patient history data alongside current examination findings requires minimal additional effort from dental care staff, and it can be seamlessly integrated with digital medical record systems. Previous research presents diverse evidence on the association between oral and cognitive health. This study builds on this knowledge by investigating different parameters and using a robust and balanced sample of participants in the MMSE 30 and MMSE  $\leq 26$  score groups.

Integrating oral health parameters into ML-driven decision support systems for MMSE score estimation could offer insights into the interplay between oral health and cognitive function. By leveraging advanced algorithms and comprehensive datasets encompassing cognitive and oral health data, researchers could uncover novel biomarkers and risk factors for cognitive decline, paving the way for more personalized approaches to cognitive assessment and intervention.

This study aims to explore the potential of using ML technologies using oral health and demographic examination data to predict the probability of having MMSE scores of 30 or  $\leq 26$  in Swedish individuals older than 60 years.

The main contributions of this study are as follows:

- The evaluation of the potential of oral health parameters for binary classification of MMSE scores of 30 and  $\leq 26$ .
- The evaluation of the most deterministic oral health parameters that influence the outcome of the best-performing ML classifier.
- The assessment of RF, SVM, and CatBoost (CB) ML classifiers as indicators of MMSE scores of 30 and  $\leq 26$ .

## Methods

### *Ethical Considerations*

This study used secondary, open-access data from the Swedish National Study on Aging and Care (SNAC-B) and Support Monitoring and Reminder Technology for Mild Dementia (SMART4MD) projects. No new data were collected specifically for this study. The use of SNAC-B data was approved by the Ethics Committee of Lund University (LU 604-00). The data for SMART4MD was collected and used under the approval of the Ethical Review Board in Sweden (LU No. 650-00 and No. 744-00). Informed consent was obtained from all participants in both studies. All data were fully anonymized prior to access by the authors. No compensation was provided to participants for the purpose of this study. All data were used in accordance with applicable data sharing agreements and stored securely. The study complies with the ethical standards outlined in the Declaration of Helsinki.

### *Data Description*

The formed sample was used in experiments with ML models, including RF, SVM, and CB, aiming to indicate whether the participant's MMSE scores were 30 or  $\leq 26$ . Data from 2 studies, which recruited participants from the European collaborative study SMART4MD [25] and the SNAC-B [26], formed the sample. SNAC-B is a longitudinal study involving individuals aged 60 years or older. Participants undergo examinations every 6 years, with more frequent assessments every 3 years, starting at the age of 78 years. The cohort ranges from 60 to 96 years of age [26]. The study sample recruited from SMART4MD involved individuals aged 60 years or older, examined every 6 to 12 months over 36 months [27]. Both studies are conducted at a research clinic affiliated with Blekinge Institute of

Technology by experienced dentists and dental hygienists and involve comprehensive clinical and demographic data collection. Ethical approval was obtained for both studies, and informed consent was collected from all participants. While both studies are longitudinal in design, only the baseline visit data, meaning the first available observation per participant, were used in this study.

All clinical and demographic features, including oral health variables, were extracted from each participant's first examination. This resulted in a cross-sectional dataset suitable for ML analysis. The final sample included 693 baseline observations (350/693, 51% females and 343/693, 49% males) with an average age of 75 years. The sample was divided into 2 relatively balanced groups: one with MMSE scores of  $\leq 26$  and one with an MMSE score of 30. Observations with intermediate scores of 27-29 were excluded to ensure distinct group separation. There was a total of 339/693 (49%) observations for individuals with  $\leq 26$  MMSE scores (165/339, 49% females and 174/339, 51% males) and 354/693 (51%) observations for the MMSE 30 score (185/354, 52% females and 169/354, 48% males). The dataset comprised 693 observations, each with 16 features (Table 1). The oral health features selected for inclusion were limited to those available in both datasets and supported by prior studies linking them to general health or cognitive decline. Examples include the Plaque Index (PI), which reflects hygiene maintenance; dry mouth, often associated with medication use and systemic conditions; and the number of teeth, which may serve as a proxy for long-term oral care and functional ability.

The listed features represent a combination of demographic, behavioral, and clinical oral health indicators obtained from baseline dental examinations. Probing pocket depth (PPD) measurements were categorized based on severity: values of 4 mm typically reflect a mild periodontal status, 5 mm indicate moderate depth, and  $\geq 6$  mm signify advanced periodontal disease. Variables such as PPD, PI, and bleeding on probing (BOP) were calculated as the percentage of affected surfaces out of the total examined. Self-reported variables, including dry mouth and hygiene tool use, were assessed through standardized questionnaires. Denture status was recorded separately for full and partial prostheses in both the maxillary and mandibular arches. These variables were selected due to their availability in both datasets and their clinical relevance to cognitive and systemic health in older adults.

MMSE scores are known to vary within individuals over time. Research indicates that changes of up to 3 points may occur without representing a clinically meaningful cognitive change, especially in older populations [28]. To reduce misclassification due to such variability, only individuals with MMSE scores of  $\leq 26$  or exactly 30 were included in the dataset. Based on this analogy, the MMSE  $\leq 26$  and the MMSE 30 score groups were designated and labeled as "1" and "0," respectively. No normalization, mean-centering, or outlier filtering was performed prior to model training. Tree-based models like RF and CB are scale-invariant, and for consistency, unscaled features were also used with SVM.

**Table 1.** List of features used for the study.

Parameter	Description	Index/Categories
Age	Participant age	• Number
Education	Level of education	• Elementary School • Secondary School • Higher education
PI <sup>a</sup>	Number of tooth surfaces in % with dental plaque of the total number of surfaces	• Percentage
PPD <sup>b</sup> 4mm	The number of PPD 4 mm in % of the total number of surfaces	• Percentage
Use of dental hygiene tools for approximal cleaning	How often do you use approximal dental hygiene tools?	• Rarely/never • Once a week • Daily • Several times daily
Gender	Gender	• Male, female
A feeling of dry mouth	Do you experience a sensation of dry mouth?	• Yes, often • Yes sometimes • No never
Number of teeth	Number of teeth	• Number
PPD 5mm	The number of PPD 5 mm in % of the total number of surfaces	• Percentage
Mirror test	Mirror test	• Glides easily • Sliding • Getting stuck
BOP <sup>c</sup>	The number of surfaces with BOP in % of the total number of surfaces.	• Percentage
PPD $\geq$ 6mm	The number of PPD $\geq$ 6 mm in % of the total number of surfaces	• Percentage
Mandibular denture	Mandibular denture	• Yes, No
Maxillary denture	Maxillary denture	• Yes, No
Mandibular partial denture	Mandibular partial denture	• Yes, No
Maxillary partial denture	Maxillary partial denture	• Yes, No

<sup>a</sup>PI: Plaque Index.

<sup>b</sup>PPD: probing pocket depth.

<sup>c</sup>BOP: bleeding on probing.

## Classification Experiment

Several well-known classifiers, RF, CB, and SVM, were used for the classification experiment. The dataset was split into 2 subgroups: a training set comprising 80% of the data (554/693 samples) and a test set containing 20% (139/693 samples). To ensure an unbiased estimation of model performance and to reduce the risk of overfitting during hyperparameter tuning, nested cross-validation (nCV) was applied within the training data.

The parameter grid for hyperparameter tuning was defined as follows:

- For Random Forest: {'n\_estimators': [50, 100, 200], 'max\_depth': [None, 10, 20], 'min\_samples\_split': [2,5,10]}
- For SVM: {'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf', 'poly'], 'degree': [2, 3, 4]}
- For CatBoost: {'iterations': [100, 200, 300], 'learning\_rate': [0.01, 0.1, 0.2], 'depth': [4, 6, 8], 'l2\_leaf\_reg': [1, 3, 5]}

nCV is particularly appropriate in ML pipelines that involve hyperparameter optimization or model selection. It separates the parameter tuning step from the performance evaluation step, which prevents information leakage between training and test data and avoids optimistically biased performance

estimates. This methodological rigor is especially important in moderate-sized datasets where variance due to overfitting can significantly affect the results [29-31].

Model performance was assessed using a 10 × 10 nCV framework. The outer 10-fold loop was used to evaluate generalization performance, while the inner loop was used solely for hyperparameter optimization. This design avoids information leakage and produces unbiased performance estimates, allowing for a robust assessment of model performance under varying fold configurations and providing a stable basis for comparison across classifiers.

## Performance Evaluation

To evaluate the performance of each ML classifier, multiple metrics were used, including accuracy, precision, recall, and  $F_1$ -score. To enhance clinical interpretability, the following definitions were used: true positives (TP) represent correctly identified positive cases; false positives (FPs) refer to negative cases incorrectly classified as positive; true negatives (TN) are correctly identified negative cases; and false negatives (FN) denote positive cases that were incorrectly classified as negative. Box plots, which illustrate the distribution of accuracy scores for the respective models and additional statistical information, were used for the data visualization. Subsequently, the classifier with the highest

scores was further analyzed for its feature importance. The examination of the most essential features relied on Shapley Additive Explanation (SHAP) summary plots, providing visualizations of the influence of parameter groups on the output, organized by their importance. This importance is based on SHAP values, offering insights into features' effects on the ML model's decision-making process [32].

## Experimental Setup

The experiment was conducted on a Dell Precision 7920 MT desktop, and all models were implemented using Python (version 3.9; Python Software Foundation) with Scikit-learn (version 1.2.2) and CB (version 1.2). Visualization was supported via Seaborn (version 0.12.2).

## Results

### Experimental Results

The dataset, comprising 693 observations, each containing 16 features in its feature vector, underwent experimentation using 3 different ML models to discriminate between MMSE  $\leq 26$  and MMSE 30 score groups. Figure 1 displays the accuracy of each model for all nCV combinations. The box plot analysis reveals that the CB model outperforms the RF and SVM models in accuracy. Specifically, the CB model achieves the highest median accuracy of 0.760, the highest mean accuracy of 0.752, and the highest maximum accuracy of 0.810. Additionally, the CB model shows the smallest variability in accuracy scores, as evidenced by the lowest IQR of 0.020. The RF model demonstrates a solid performance with a median and mean accuracy of 0.740, but it does not reach the same maximum accuracy as the CB model, peaking at 0.780. The RF model also has a moderate IQR of 0.030, indicating some variability in its performance. While still competitive, the SVM model shows the lowest median accuracy of 0.730 and the lowest mean accuracy of 0.721 among the 3 models. It also has a maximum accuracy of

0.760, which is lower than both RF and CB. The IQR of 0.030 is comparable to that of RF, indicating a similar level of performance variability.

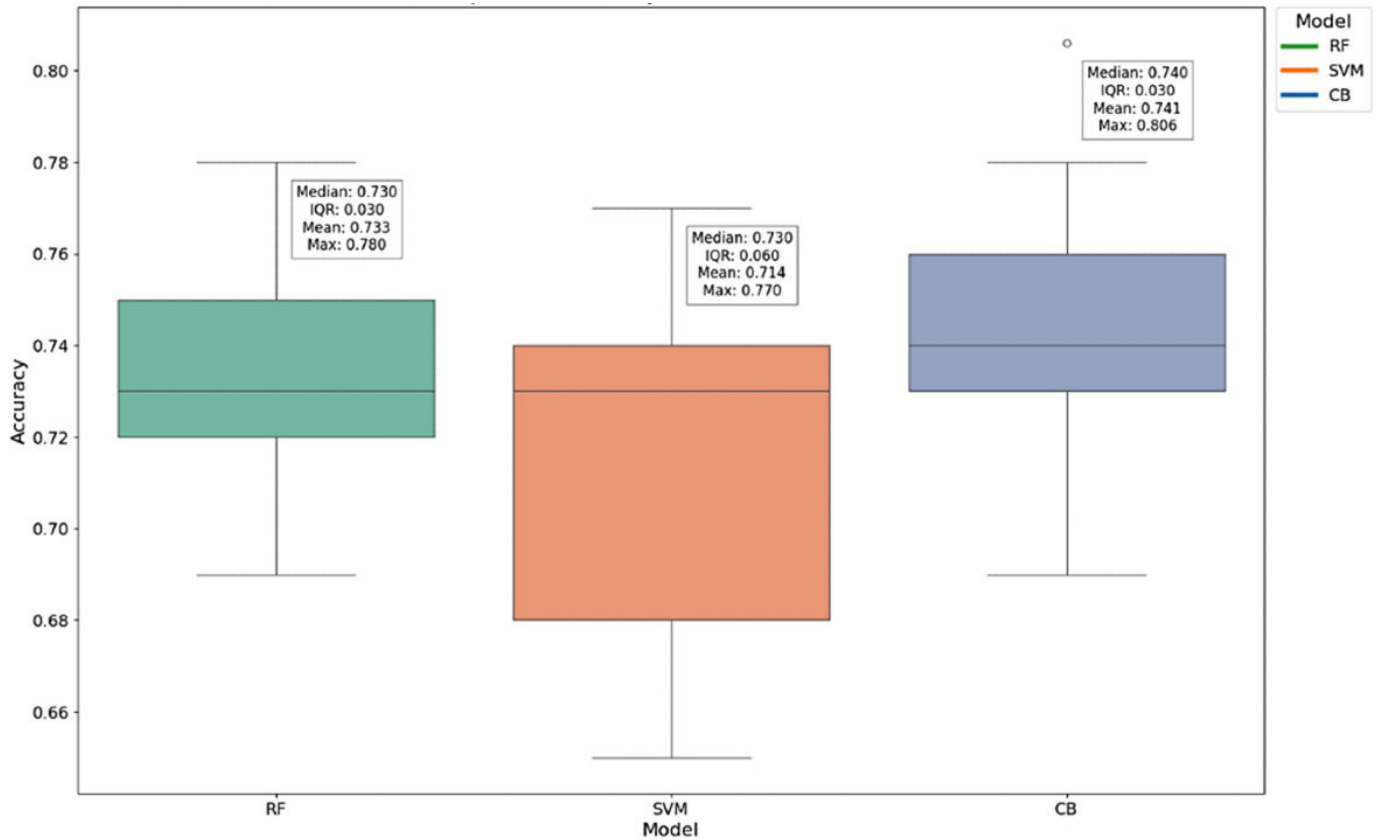
The performance of the 3 ML models (CB, RF, and SVM) was evaluated using accuracy, precision, recall, and  $F_1$ -score on both the training and test datasets. Table 2 summarizes these results. On the training set, RF achieved the highest scores across all metrics, followed closely by CB. However, on the test set, CB demonstrated the most balanced and generalizable performance, with an accuracy of 80.6%, precision of 77.4%, recall of 78.7%, and an  $F_1$ -score of 78.0%. In contrast, RF showed signs of overfitting, indicated by a decline in test precision and accuracy despite maintaining high recall. SVM yielded moderate but consistent results across both datasets. These results suggest that CB is better suited for generalization in this context, likely due to its regularization features. RF, while highly effective on the training set, may have captured dataset-specific patterns that did not generalize well. SVM demonstrated stable but lower overall performance.

Table 3 shows the confusion matrix results associated with all models. The performance of 3 ML models, CB, RF, and SVM, was evaluated on both training and test datasets in terms of TP, FP, TN, and FN. Rather than focusing on raw counts, the interpretation emphasizes overall performance patterns. CB maintained a strong balance between correctly identifying both positive and negative cases. RF achieved high recall but also produced more FPs on the test set, which may limit its clinical precision. SVM showed more balanced errors, though with generally lower performance.

The hyperparameters and nCV combinations associated with each and the best-performing model are as follows:

- CB: 3X3nCV, ('depth': 4, 'iterations': 200, 'l2\_leaf\_reg': 5, 'learning rate': 0.1).
- RF: 4X7nCV, ('max depth': 20, 'min samples split': 5, 'n estimators': 50).
- SVM: 8X5nCV, ('C': 10, 'degree': 3, 'kernel': 'poly').

**Figure 1.** Box plot-based accuracy comparison of each machine learning model. CB: CatBoost; RF: random forests; SVM: support vector machine.



**Table 2.** The average performance metric scores for each machine learning model are displayed for the training and test sets.

Model	Precision (%)	Recall (%)	Accuracy (%)	F <sub>1</sub> -score (%)
Training set				
CB <sup>a</sup>	90.2	87.4	88.6	88.8
RF <sup>b</sup>	90.4 <sup>c</sup>	92.6 <sup>c</sup>	91.2 <sup>c</sup>	91.5 <sup>c</sup>
SVM <sup>d</sup>	75.8	69.1	72.7	72.3
Test set				
CB	77.4 <sup>c</sup>	78.7	80.6 <sup>c</sup>	78.0 <sup>c</sup>
RF	68.5	82.0 <sup>c</sup>	75.6	74.6
SVM	75.3	68.4	72.2	71.7

<sup>a</sup>CB: CatBoost.

<sup>b</sup>RF: random forests.

<sup>c</sup>Top-performing values.

<sup>d</sup>SVM: support vector machine.

**Table 3.** The confusion matrix outcomes for every machine learning classifier on both the validation and test sets.

	CB <sup>a</sup>		RF <sup>b</sup> Predictedd)		SVM <sup>c</sup>	
	+	-	+	-	+	-
Training set						
Actual						
Positive (+)	249	36	264	21	195	90
Negative (-)	27	242	28	241	64	205
Test set						
Actual						
Positive (+)	48	13	50	11	46	15
Negative (-)	14	64	23	55	17	61

	CB <sup>a</sup>		RF <sup>b</sup> Predicted(d)		SVM <sup>c</sup>	
	+	-	+	-	+	-

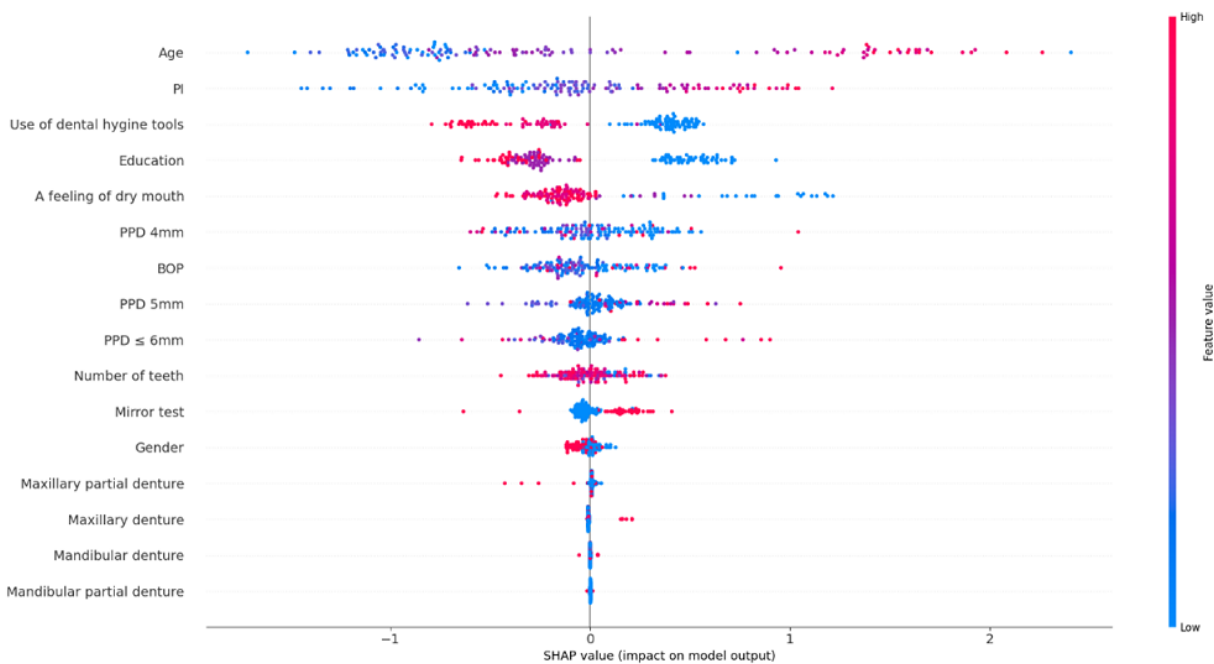
<sup>a</sup>CB: CatBoost.  
<sup>b</sup>RF: random forests.  
<sup>c</sup>SVM: support vector machine.

## Feature Importance

The SHAP summary plot in Figure 2 illustrates the impact of various features on the model's output. Each dot represents a Shapley value for a feature for an individual instance, and the color denotes the feature value, ranging from low (blue) to high (red). The SHAP summary plot reveals that age, PI, dental hygiene devices, education, a feeling of dry mouth, and PPD are critical in influencing the model's predictions. The balance of these features, along with others like BOP and the number of teeth, determines the overall outcome predicted by the model. These findings indicate that age, PI, and PPD are

significant predictors, with older age and higher PI correlating with an increased likelihood of having an MMSE score  $\leq 26$ . Education appears to inversely affect the prediction, suggesting that higher education levels may be associated with better oral health practices. The number of teeth shows an inverse relationship, where having more teeth is linked to a lower risk of MMSE  $\leq 26$ . Additionally, the presence of dental prosthetics and feelings of dry mouth positively influence the prediction, indicating a higher likelihood of MMSE  $\leq 26$ . Gender shows a minimal impact, with the female gender slightly reducing the predicted MMSE  $\leq 26$  risk.

**Figure 2.** SHAP summary plot for model features. The plot illustrates the impact of each feature on the model's output. BOP: bleeding on probing; PI: Plaque Index; PPD: probing pocket depth; SHAP: Shapley Additive Explanation.



## Discussion

### Principal Findings

This study explored the potential of using ML technologies using oral health and demographic examination data to predict the probability of having MMSE scores of 30 or  $\leq 26$  in Swedish individuals older than 60 years. The main findings of the study were that all CB, RF, and SVM classifiers achieved high classification accuracies, but the CB classifier outperformed both RF and SVM classifiers with an average accuracy of 80.6%. Acceptable accuracy levels vary across studies [33-35], but it can be argued that all CB, RF, and SVM classifiers demonstrate strong performance, each exceeding 72% accuracy. This indicates that with specific

nCV combinations and hyperparameter tuning, all classifiers can indicate whether the individual belongs to  $\leq 26$  MMSE scores or higher MMSE score (score=30) groups with varying grades of accuracy [36]. These findings hold the potential to identify individuals in need of support and treatment, benefiting not only them but also health care professionals and systems by optimizing treatment with fewer resources.

The performance evaluation based on the training and test set results is a common indicator of finding the unexpected accuracy relation of ML models [37]. The evaluation of the models reveals distinct differences in performance and generalization capabilities. On the training set, RF showed a well-balanced performance with strong metrics across the board, reflected in its Precision, Recall, Accuracy, and  $F_1$ -score and its TP, TN, FP, and FN counts. The classifier

CB performs better in the test set regarding the average performance metric, and SVM lagged significantly behind in both training and test sets. However, the general performance evaluation might differ from a clinical point of view when the higher recall is considered a more important parameter. In that case, RF with the highest TP values becomes the best classifier. FP could lead to unnecessary clinical evaluations or interventions, while FN could result in missed diagnoses, delaying treatment for cognitive impairment. When the goal of ML models is early detection and ensuring no cases of cognitive impairment are missed, minimizing FN is essential [38]. However, if the focus of these models is on reducing unnecessary interventions and managing health care resources efficiently, minimizing FP should be prioritized [39]. From a generalizability perspective, SVM shows the lowest performance drop between training and test, and CB indicates a second strong generalizability with a low-performance drop compared with the training and test set results. On the other hand, RF showed a decline in performance between the training and test sets. This suggests that SVM and CB are more suitable than RF for handling new, unseen data. Hyperparameter tuning was conducted through nCV using model-specific grids. For RF, parameters included the number of estimators, tree depth, and minimum samples per split. Despite this rigorous tuning, RF exhibited overfitting, reflected by its high training accuracy and reduced performance on the test set. This behavior suggests that RF, with its tendency to create deep and complex trees, may have captured noise or sample-specific patterns. In contrast, CB's built-in regularization mechanisms, including L2-leaf regularization, contributed to more stable generalization. To address overfitting in future implementations, simpler models or additional regularization techniques such as pruning, reduced tree depth, or feature selection could be explored.

Oral health and demographic factors vary across different regions globally, and this method's applicability is specific to Swedish conditions, which poses a limitation. However, a study from Taiwan [24], using oral health biomarkers and CB for cognitive decline classification predictions, yielded results akin to those presented in this study. In their model [24], crucial features included daily social interaction, oral diadochokinesis, PI, saliva protein, and age. In contrast, this study benefits from a more extensive and balanced dataset, presenting absolute average values rather than weighted results. Additionally, to mitigate overfitting risks, nCV was used to train ML classifiers across various k-fold combinations to mitigate overfitting risks in ML classifiers across various k-fold combinations. Furthermore, these studies collectively highlight oral health and demographic factors that vary across different regions globally, and this method's applicability is specific to Swedish conditions. Additionally, while the use of  $10 \times 10$  nCV ensures an unbiased estimate of generalization performance within the dataset, the importance of validating models on external data is well recognized. Internal validation alone cannot capture the variability introduced by different clinical environments, geographic regions, or data acquisition protocols. External validation on independent cohorts or through integration into routine dental examinations would be required to evaluate robustness. Such

efforts would help assess the model's applicability under real-world variability and determine its readiness for broader clinical use.

The implications drawn from the SHAP summary plot underscore critical insights into feature relevance concerning cognitive health assessments via ML models. The clear demarcation between the groups of MMSE scores  $\leq 26$  and 30 across influential variables like age suggests that such demographic factors hold substantial weight in predictive analytics. Notably, age emerges as a pivotal factor, with higher age correlating strongly with lower cognitive health scores. This correlation is also highlighted in other studies that consistently demonstrated that MMSE scores tend to be lower for older individuals and decline over time [40]. The SHAP summary plot highlights age, PI, and use of dental hygiene tools as the most influential predictors of MMSE classification. Older age and higher plaque levels were associated with increased likelihood of MMSE  $\leq 26$ , suggesting that both biological aging and reduced oral hygiene maintenance contribute to cognitive risk [20,41]. In particular, individuals with lower use of interdental hygiene tools exhibited stronger SHAP effects, indicating a potential interaction between age-related decline in motor or cognitive function and oral hygiene behaviors. This aligns with findings that individuals with cognitive impairment often experience challenges in maintaining oral hygiene due to reduced executive functioning and manual dexterity [7,42]. Features such as dry mouth and education also showed moderate predictive strength, consistent with their known associations with polypharmacy and cognitive reserve, respectively [41]. In contrast, features like BOP, gender, and presence of dentures showed minimal SHAP influence, suggesting weaker or less consistent relationships within this sample. These findings suggest that age-related decline may act synergistically with deteriorating oral hygiene habits, offering further evidence for behaviorally anchored early risk screening in clinical practice.

Given that MCI resides on the border between healthy aging and neurocognitive illness, individuals with MCI may fluctuate between MMSE scores  $\leq 26$  and 30 [43]. About 50% of individuals with MCI receive a neurocognitive diagnosis within 3 years, with a 2-step decrease in MMSE score within 12 months indicating disease onset. Recommended thresholds for the MMSE can be expressed as 0-22, indicating dementia; 23-26, MCI [44]; and 27-30, indicating normal cognitive function [7,42]. In this study, the threshold MMSE  $\leq 26$  was used. This criterion was adopted in the study, resulting in the exclusion of MMSE scores of 27, 28, and 29 from calculations. This limit has been chosen as 2 different studies have found that changes of 2 to 3 points indicate reliable changes in the MMSE test value at the 90% confidence level [28,45]. However, it is important to note that this regulatory threshold may limit the study, potentially leading to missed indications. The exclusion of MMSE scores of 27, 28, and 29 may also affect generalizability. One of the strengths of this study is the use of ML as the methodology. Recent studies in ML have shown significant promise in predicting various disorders, such as dementia [46,47], chronic obstructive pulmonary



disease [48], cancer [49], and Parkinson disease [50]. The used tree-based models, CB and RF, and the distance-based model SVM allow comparability in the case of possible effects of outliers in the dataset, where the tree-based models are known to be more resistant to imbalance in the dataset than the distance-based models [51-53]. Despite this study seeking a balanced data set, there is a heterogeneity regarding ages represented in the dataset where some ages might be overrepresented, which can be seen as a limitation in this study. Another limitation is the low number of participants and the recruitment that took place in a small county in southern Sweden, which is an obstacle to making conclusions from a broader perspective that represents more variety in demographic and regional differences. However, in relation to previous research, this study includes a larger number of participants, thereby increasing the generalizability of the findings. Another notable difference concerns the threshold used to distinguish between the MMSE  $\leq 26$  and MMSE 30 groups; in this study, a higher threshold was applied. Since the dataset of 693 observations is relatively small and may not be representative of a larger population, it may affect generalizability. In addition, even if the groups are relatively balanced, small differences in group size (339/693, 49% vs 354/693, 51%) can affect model performance and bias. These limitations can be overcome by expanding the data set by including divergent and wider participants' data from larger geographical areas in upcoming studies. Future research may also address the performance of capturing changes over time for monitoring purposes, highlighting the need for longitudinal studies. Although the findings demonstrate strong internal performance, several practical challenges constrain the model's readiness for clinical use. These include variability in oral health assessment protocols, limited interpretability of model outputs for nontechnical users, and the absence of integration pathways with existing clinical information systems. SHAP plots provide valuable transparency but may need adaptation for clinician-facing use.

Implementation would require standardization of input data, compatibility with electronic health records, and interdisciplinary development efforts. Further investigation is needed to evaluate clinical usability and workflow alignment.

Other limitations of the study are that only baseline data were used from longitudinal studies, meaning that potential changes over time may be missed. Only 16 parameters were selected, which means that other important parameters that could affect the results may have been omitted. The oral health features used in the models were selected based on their consistent availability across both datasets and their previously reported clinical relevance to cognitive function. These features, such as PI, number of teeth, and dry mouth, may reflect both localized oral conditions and systemic factors, including frailty, polypharmacy, and reduced self-care capacity. Such conditions have been associated with cognitive impairment in older adults. Other potentially relevant features, like microbial profiles or periodontal attachment levels, were not included due to data limitations, but future work could incorporate these variables in extended datasets.

## Conclusion

This study suggests that oral health parameters, demographic data, and ML techniques offer a potential tool for screening MMSE scores for individuals aged 60 years and older. The data used as inputs for ML classifiers contains sufficient information to differentiate between MMSE scores  $\leq 26$  and 30.

It is important to include both demographic and oral health-related factors in predictive models. Despite some limitations, the study offers valuable insights for future research and potential clinical applications. To ensure the robustness and generalizability of the model, further external validation and adaptation to different clinical settings are required.

---

## Acknowledgments

We want to express our greatest appreciation to the participants and staff of the Swedish National Study on Aging and Care and Support Monitoring and Reminder Technology for Mild Dementia studies. The Blekinge Institute of Technology funded this study.

---

## Data Availability

The datasets analyzed during this study are not publicly available due to licensing restrictions from Blekinge Institute of Technology, Department of Health, but are available from the corresponding author upon reasonable request and with permission from Blekinge Institute of Technology. The Python code can be found on [54].

---

## Author Contributions

AI and JF conducted the experiment and analyzed the data. AI and JF drafted the paper, and then SNG, EMM, HD, AA, and ALD critically revised the text regarding intellectually important content. All authors have given final approval for publication and take public responsibility for all its parts.

---

## Conflicts of Interest

None declared.

## References

1. World population prospects 2024. United Nations Population Division. URL: <https://population.un.org/wpp/> [Accessed 2025-04-02]
2. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers Dement*. Jan 2013;9(1):63. [doi: [10.1016/j.jalz.2012.11.007](https://doi.org/10.1016/j.jalz.2012.11.007)]
3. Yiannopoulou KG, Papageorgiou SG. Current and future treatments in Alzheimer disease: an update. *J Cent Nerv Syst Dis*. 2020;12:1179573520907397. [doi: [10.1177/1179573520907397](https://doi.org/10.1177/1179573520907397)] [Medline: [32165850](https://pubmed.ncbi.nlm.nih.gov/32165850/)]
4. Long JM, Holtzman DM. Alzheimer disease: an update on pathobiology and treatment strategies. *Cell*. Oct 3, 2019;179(2):312-339. [doi: [10.1016/j.cell.2019.09.001](https://doi.org/10.1016/j.cell.2019.09.001)] [Medline: [31564456](https://pubmed.ncbi.nlm.nih.gov/31564456/)]
5. Park KW. Anti-amyloid antibody therapies for Alzheimer's disease. *Nucl Med Mol Imaging*. Jun 2024;58(4):227-236. [doi: [10.1007/s13139-024-00848-3](https://doi.org/10.1007/s13139-024-00848-3)] [Medline: [38932758](https://pubmed.ncbi.nlm.nih.gov/38932758/)]
6. Syrjälä AH, Ylöstalo P, Ruoppi P, et al. Dementia and oral health among subjects aged 75 years or older. *Gerodontology*. Mar 2012;29(1):36-42. [doi: [10.1111/j.1741-2358.2010.00396.x](https://doi.org/10.1111/j.1741-2358.2010.00396.x)]
7. Gil-Montoya JA, Sánchez-Lara I, Carnero-Pardo C, et al. Oral hygiene in the elderly with different degrees of cognitive impairment and dementia. *J Am Geriatr Soc*. Mar 2017;65(3):642-647. [doi: [10.1111/jgs.14697](https://doi.org/10.1111/jgs.14697)] [Medline: [28024093](https://pubmed.ncbi.nlm.nih.gov/28024093/)]
8. Flyborg J, Renvert S, Sanmartin Berglund J, Anderberg P. Use of a powered toothbrush to improve oral health in individuals with mild cognitive impairment. *Gerodontology*. Mar 2023;40(1):74-82. [doi: [10.1111/ger.12619](https://doi.org/10.1111/ger.12619)] [Medline: [35064682](https://pubmed.ncbi.nlm.nih.gov/35064682/)]
9. Guo H, Chang S, Pi X, et al. The effect of periodontitis on dementia and cognitive impairment: a meta-analysis. *Int J Environ Res Public Health*. Jun 25, 2021;18(13):34202071. [doi: [10.3390/ijerph18136823](https://doi.org/10.3390/ijerph18136823)] [Medline: [34202071](https://pubmed.ncbi.nlm.nih.gov/34202071/)]
10. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. Nov 1975;12(3):189-198. [doi: [10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)] [Medline: [1202204](https://pubmed.ncbi.nlm.nih.gov/1202204/)]
11. Roshanzamir A, Aghajan H, Soleymani Baghshah M. Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Med Inform Decis Mak*. Mar 9, 2021;21(1):92. [doi: [10.1186/s12911-021-01456-3](https://doi.org/10.1186/s12911-021-01456-3)] [Medline: [33750385](https://pubmed.ncbi.nlm.nih.gov/33750385/)]
12. Parsapoor Parsa MM, Alam MR, Mihailidis A. Performance of machine learning algorithms for dementia assessment: impacts of language tasks, recording media, and modalities. *BMC Med Inform Decis Mak*. Mar 3, 2023;23(1):45. [doi: [10.1186/s12911-023-02122-6](https://doi.org/10.1186/s12911-023-02122-6)] [Medline: [36869377](https://pubmed.ncbi.nlm.nih.gov/36869377/)]
13. Vyas A, Aisopos F, Vidal ME, Garrard P, Paliouras G. Identifying the presence and severity of dementia by applying interpretable machine learning techniques on structured clinical records. *BMC Med Inform Decis Mak*. Oct 17, 2022;22(1):271. [doi: [10.1186/s12911-022-02004-3](https://doi.org/10.1186/s12911-022-02004-3)] [Medline: [36253849](https://pubmed.ncbi.nlm.nih.gov/36253849/)]
14. Kang MJ, Kim SY, Na DL, et al. Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. *BMC Med Inform Decis Mak*. Nov 21, 2019;19(1):231. [doi: [10.1186/s12911-019-0974-x](https://doi.org/10.1186/s12911-019-0974-x)] [Medline: [31752864](https://pubmed.ncbi.nlm.nih.gov/31752864/)]
15. Twait EL, Andaur Navarro CL, Gudnason V, Hu YH, Launer LJ, Geerlings MI. Dementia prediction in the general population using clinically accessible variables: a proof-of-concept study using machine learning. the AGES-Reykjavik study. *BMC Med Inform Decis Mak*. Aug 28, 2023;23(1):168. [doi: [10.1186/s12911-023-02244-x](https://doi.org/10.1186/s12911-023-02244-x)] [Medline: [37641038](https://pubmed.ncbi.nlm.nih.gov/37641038/)]
16. Luo S, Yao X, Yuan Z, Zhou L. Prediction of mini-mental state examination scores via machine learning for Alzheimer's disease. In: You P, Li H, Chen Z, editors. *Proceedings of International Conference on Image, Vision and Intelligent Systems 2022 (ICIVIS 2022)*. Springer Nature; 2023:878-883. [doi: [10.1007/978-981-99-0923-0\\_87](https://doi.org/10.1007/978-981-99-0923-0_87)]
17. Wang J, Wang Z, Liu N, et al. Random forest model in the diagnosis of dementia patients with normal mini-mental state examination scores. *J Pers Med*. Jan 4, 2022;12(1):37. [doi: [10.3390/jpm12010037](https://doi.org/10.3390/jpm12010037)] [Medline: [35055352](https://pubmed.ncbi.nlm.nih.gov/35055352/)]
18. Qiu S, Chang GH, Panagia M, Gopal DM, Au R, Kolachalama VB. Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alz & Dem Diag Ass & Dis Mo*. Jan 2018;10(1):737-749. [doi: [10.1016/j.dadm.2018.08.013](https://doi.org/10.1016/j.dadm.2018.08.013)]
19. Zeng LN, Zong QQ, Xu SW, et al. Oral health in patients with dementia: a meta-analysis of comparative and observational studies. *Int J Geriatr Psychiatry*. Apr 2021;36(4):467-478. [doi: [10.1002/gps.5453](https://doi.org/10.1002/gps.5453)] [Medline: [33105039](https://pubmed.ncbi.nlm.nih.gov/33105039/)]
20. Hamza SA, Asif S, Bokhari SAH. Oral health of individuals with dementia and Alzheimer's disease: a review. *J Indian Soc Periodontol*. 2021;25(2):96-101. [doi: [10.4103/jisp.jisp\\_287\\_20](https://doi.org/10.4103/jisp.jisp_287_20)] [Medline: [33888939](https://pubmed.ncbi.nlm.nih.gov/33888939/)]
21. Wei T, Du Y, Hou T, et al. Association between adverse oral conditions and cognitive impairment: a literature review. *Front Public Health*. 2023;11:1147026. [doi: [10.3389/fpubh.2023.1147026](https://doi.org/10.3389/fpubh.2023.1147026)] [Medline: [37089515](https://pubmed.ncbi.nlm.nih.gov/37089515/)]
22. Inamochi Y, Ogino Y, Harada K, et al. Do oral care and rehabilitation improve cognitive function? a systematic review of clinical studies. *J Evid Based Dent Pract*. Mar 2024;24(1):101948. [doi: [10.1016/j.jebdp.2023.101948](https://doi.org/10.1016/j.jebdp.2023.101948)]

23. Zuluaga DJM, Montoya JAG, Contreras CI, Herrera RR. Association between oral health, cognitive impairment and oral health-related quality of life. *Gerodontology*. Jun 2012;29(2):e667-73. [doi: [10.1111/j.1741-2358.2011.00542.x](https://doi.org/10.1111/j.1741-2358.2011.00542.x)] [Medline: [21883423](https://pubmed.ncbi.nlm.nih.gov/21883423/)]
24. Chung PC, Chan TC. Digital oral health biomarkers for early detection of cognitive decline. *BMC Public Health*. Oct 9, 2023;23(1):1952. [doi: [10.1186/s12889-023-16897-w](https://doi.org/10.1186/s12889-023-16897-w)] [Medline: [37814231](https://pubmed.ncbi.nlm.nih.gov/37814231/)]
25. Support, monitoring and reminder technology for mild dementia. SMART4MD project. European Commission. URL: <https://cordis.europa.eu/project/id/643399> [Accessed 2021-10-01]
26. Lagergren M, Fratiglioni L, Hallberg IR, et al. A longitudinal study integrating population, care and social services data. the Swedish national study on Aging and Care (SNAC). *Aging Clin Exp Res*. Apr 2004;16(2):158-168. [doi: [10.1007/BF03324546](https://doi.org/10.1007/BF03324546)] [Medline: [15195992](https://pubmed.ncbi.nlm.nih.gov/15195992/)]
27. Flyborg J, Renvert S, Anderberg P, Sanmartin-Berglund J. The long-term effect on oral health and quality of life using a powered toothbrush in individuals with mild cognitive impairment. an intervention trial. *Spec Care Dentist*. 2024;44(6):1700-1708. [doi: [10.1111/scd.13040](https://doi.org/10.1111/scd.13040)] [Medline: [38994574](https://pubmed.ncbi.nlm.nih.gov/38994574/)]
28. Stein J, Luppa M, Maier W, et al. Assessing cognitive changes in the elderly: reliable change indices for the mini-mental state examination. *Acta Psychiatr Scand*. Sep 2012;126(3):208-218. [doi: [10.1111/j.1600-0447.2012.01850.x](https://doi.org/10.1111/j.1600-0447.2012.01850.x)] [Medline: [22375927](https://pubmed.ncbi.nlm.nih.gov/22375927/)]
29. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. Feb 23, 2006;7(1):91. [doi: [10.1186/1471-2105-7-91](https://doi.org/10.1186/1471-2105-7-91)] [Medline: [16504092](https://pubmed.ncbi.nlm.nih.gov/16504092/)]
30. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. Aug 1, 2010;11:2079-2107. URL: <https://www.jmlr.org/papers/volume11/cawley10a/cawley10a.pdf> [Accessed 2025-08-21]
31. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Statist Surv*. Jan 2010;4(none):40-79. [doi: [10.1214/09-SS054](https://doi.org/10.1214/09-SS054)]
32. Marcilio WE, Eler DM. From explanations to feature selection: assessing SHAP values as feature selection mechanism. Presented at: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); Nov 7-10, 2020:340-347; Porto de Galinhas, Brazil. 2020.[doi: [10.1109/SIBGRAPI51738.2020.00053](https://doi.org/10.1109/SIBGRAPI51738.2020.00053)]
33. Cabitza F, Campagner A, Del Zotti F, Verona N, Ravizza A, Sternini F. All you need is higher accuracy? on the quest for minimum acceptable accuracy for medical artificial intelligence. Presented at: E-Health Proceedings 12th International Conference on e-Health; Sep 23-25, 2020. 2020.
34. Ebrahim M, Sedky AAH, Mesbah S. Accuracy assessment of machine learning algorithms used to predict breast cancer. *Data (Basel)*. Feb 2023;8(2):35. [doi: [10.3390/data8020035](https://doi.org/10.3390/data8020035)]
35. Agliata A, Giordano D, Bardozzo F, Bottiglieri S, Facchiano A, Tagliaferri R. Machine learning as a support for the diagnosis of type 2 diabetes. *Int J Mol Sci*. Apr 5, 2023;24(7):6775. [doi: [10.3390/ijms24076775](https://doi.org/10.3390/ijms24076775)] [Medline: [37047748](https://pubmed.ncbi.nlm.nih.gov/37047748/)]
36. How to know if your machine learning model has good performance. Obviously AI. URL: <https://www.obviously.ai/post/machine-learning-model-performance> [Accessed 2024-05-29]
37. Yoon H. Finding unexpected test accuracy by cross validation in machine learning. *Int J Comput Sci Netw Secur*. 2021;21(12spc):549-555. [doi: [10.22937/IJCSNS.2021.21.12.76](https://doi.org/10.22937/IJCSNS.2021.21.12.76)]
38. Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K. False-negative results in screening programmes: systematic review of impact and implications. *Health Technol Assess*. 2000;4(5). [doi: [10.3310/hta4050](https://doi.org/10.3310/hta4050)]
39. Dresselhaus TR, Luck J, Peabody JW. The ethical problem of false positives: a prospective evaluation of physician reporting in the medical record. *J Med Ethics*. Oct 2002;28(5):291-294. [doi: [10.1136/jme.28.5.291](https://doi.org/10.1136/jme.28.5.291)] [Medline: [12356955](https://pubmed.ncbi.nlm.nih.gov/12356955/)]
40. Piccinin AM, Muniz-Terrera G, Clouston S, et al. Coordinated analysis of age, sex, and education effects on change in MMSE scores. *J Gerontol B Psychol Sci Soc Sci*. May 1, 2013;68(3):374-390. [doi: [10.1093/geronb/gbs077](https://doi.org/10.1093/geronb/gbs077)]
41. Ghezzi EM, Ship JA. Dementia and oral health. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. Jan 2000;89(1):2-5. [doi: [10.1016/s1079-2104\(00\)80003-7](https://doi.org/10.1016/s1079-2104(00)80003-7)] [Medline: [10630932](https://pubmed.ncbi.nlm.nih.gov/10630932/)]
42. Marchini L, Ettinger R, Caprio T, Jucan A. Oral health care for patients with Alzheimer's disease: an update. *Spec Care Dentist*. May 2019;39(3):262-273. [doi: [10.1111/scd.12375](https://doi.org/10.1111/scd.12375)] [Medline: [30964560](https://pubmed.ncbi.nlm.nih.gov/30964560/)]
43. Overton M, Pihlsgård M, Elmståhl S. Diagnostic stability of mild cognitive impairment, and predictors of reversion to normal cognitive functioning. *Dement Geriatr Cogn Disord*. 2020. [doi: [10.1159/000506255](https://doi.org/10.1159/000506255)]
44. Delavaran H, Jönsson AC, Lövkvist H, et al. Cognitive function in stroke survivors: a 10-year follow-up study. *Acta Neurol Scand*. Sep 2017;136(3):187-194. [doi: [10.1111/ane.12709](https://doi.org/10.1111/ane.12709)]
45. Hensel A, Angermeyer MC, Riedel-Heller SG. Measuring cognitive change in older adults: reliable change indices for the mini-mental state examination. *J Neurol Neurosurg Psychiatry*. Dec 2007;78(12):1298-1303. [doi: [10.1136/jnnp.2006.109074](https://doi.org/10.1136/jnnp.2006.109074)] [Medline: [17442763](https://pubmed.ncbi.nlm.nih.gov/17442763/)]

46. Nyholm J, Ghazi AN, Ghazi SN, Sanmartin Berglund J. Prediction of dementia based on older adults' sleep disturbances using machine learning. *Comput Biol Med.* Mar 2024;171:108126. [doi: [10.1016/j.compbiomed.2024.108126](https://doi.org/10.1016/j.compbiomed.2024.108126)] [Medline: [38342045](https://pubmed.ncbi.nlm.nih.gov/38342045/)]
47. Javeed A, Dallora AL, Berglund JS, et al. Early prediction of dementia using feature extraction battery (FEB) and optimized support vector machine (SVM) for classification. *Biomedicines.* Feb 2, 2023;11(2):439. [doi: [10.3390/biomedicines11020439](https://doi.org/10.3390/biomedicines11020439)] [Medline: [36830975](https://pubmed.ncbi.nlm.nih.gov/36830975/)]
48. Idrisoglu A, Dallora AL, Cheddad A, Anderberg P, Jakobsson A, Sanmartin Berglund J. COPDVD: automated classification of chronic obstructive pulmonary disease on a new collected and evaluated voice dataset. *Artif Intell Med.* Oct 2024;156:102953. [doi: [10.1016/j.artmed.2024.102953](https://doi.org/10.1016/j.artmed.2024.102953)] [Medline: [39222579](https://pubmed.ncbi.nlm.nih.gov/39222579/)]
49. Cheddad A. Machine learning in healthcare: breast cancer and diabetes cases. In: Reis T, Borschlegl MX, Angelini M, Hemmje ML, editors. *Advanced Visual Interfaces Supporting Artificial Intelligence and Big Data Applications.* 2021:125-135. [doi: [10.1007/978-3-030-68007-7\\_8](https://doi.org/10.1007/978-3-030-68007-7_8)]
50. Saleh S, Cherradi B, El Gannour O, Hamida S, Bouattane O. Predicting patients with Parkinson's disease using machine learning and ensemble voting technique. *Multimed Tools Appl.* 2024;83(11):33207-33234. [doi: [10.1007/s11042-023-16881-x](https://doi.org/10.1007/s11042-023-16881-x)]
51. Breiman L. Random forests. *Mach Learn.* Oct 2001;45(1):5-32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
52. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Presented at: *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*; Dec 3-8, 2018; Montréal, Canada. 2018.URL: [https://proceedings.neurips.cc/paper\\_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html) [Accessed 2024-09-11]
53. Baldomero-Naranjo M, Martínez-Merino LI, Rodríguez-Chía AM. A robust SVM-based approach with feature selection and outliers detection for classification problems. *Expert Syst Appl.* Sep 2021;178:115017. [doi: [10.1016/j.eswa.2021.115017](https://doi.org/10.1016/j.eswa.2021.115017)]
54. AIITPlanet/code. GitHub. URL: <https://github.com/AIITPlanet/Code/blob/main/ClassificationOfMMSE.py> [Accessed 2025-06-24]

## Abbreviations

**AD:** Alzheimer Disease

**BOP:** bleeding on probing

**CB:** CatBoost

**FN:** false negative

**FP:** false positive

**MCI:** mild cognitive impairment

**ML:** machine learning

**MMSE:** Mini-Mental State Examination

**nCV:** nested cross-validation

**PI:** Plaque Index

**PPD:** probing pocket depth

**RF:** random forest

**SHAP:** Shapley additive explanation

**SMART4MD:** Support Monitoring and Reminder Technology for Mild Dementia

**SNAC-B:** Swedish National Study on Aging and Care

**SVM:** support vector machine

**TN:** true negative

**TP:** true positive

*Edited by Arriel Benis; peer-reviewed by Kudirat Abidemi Obisesan Olawuni, Soumya Kavuluri, Zhang Xiang; submitted 27.03.2025; final revised version received 24.06.2025; accepted 15.07.2025; published 25.08.2025*

*Please cite as:*

*Idrisoglu A, Flyborg J, Nauman Ghazi S, Mikaelsson Midlöv E, Dellkvist H, Axén A, Dallora AL*

*Prediction of Mini-Mental State Examination Scores for Cognitive Impairment and Machine Learning Analysis of Oral Health and Demographic Data Among Individuals Older Than 60 Years: Cross-Sectional Study*

*JMIR Med Inform 2025;13:e75069*

*URL: <https://medinform.jmir.org/2025/1/e75069>*

*doi: [10.2196/75069](https://doi.org/10.2196/75069)*

© Alper Idrisoglu, Johan Flyborg, Sarah Nauman Ghazi, Elina Mikaelsson Midlöv, Helén Dellkvist, Anna Axén, Ana Luiza Dallora. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 25.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.