Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 12(1): 214-223 (2024)

A Novel Approach to Detection of Alzheimer's Disease from Handwriting: Triple Ensemble Learning Model

Hakan ÖCAL[1*]

[1]Bartın University, Faculty of Engineering Architecture and Design, Department of Computer Engineering, Bartın, Turkey

Article Info

Research article
Received: 05/10/2023
Revision: 12/02/2024
Accepted: 13/02/2024

Keywords

Neurodegenerative disease
Alzheimer's disease prediction
Ensemble machine learning model
Classification
Handwriting dataset

Makale Bilgisi

Araştırma makalesi
Başvuru: 05/10/2023
Düzeltme: 12/02/2024
Kabul: 13/02/2024

Anahtar Kelimeler

Nörodejeneratif hastalık
Alzheimer hastalığı tahmini
Topluluk makine öğrenimi modeli
Sınıflandırma
El yazısı veri kümeleri

Graphical/Tabular Abstract (Grafik Özet)

To detect Alzheimer's disease from handwriting, Light Gradient Boosting Machine, Categorical Boosting, and Adaptive Boosting machine learning classification algorithms were combined with a Hard Voting Classifier and trained and tested on the publicly available Diagnosis Alzheimer's With haNdwriting dataset. / Alzheimer hastalığını el yazısından tespit etmek için Gradient Boosting Machine, Kategorik Boosting ve Adaptive Boosting makine öğrenimi sınıflandırma algoritmaları, Hard Voting Classifier ile birleştirildi ve halka açık Diagnosis Alzheimer's with haNdwriting veri kümesi üzerinde eğitildi.

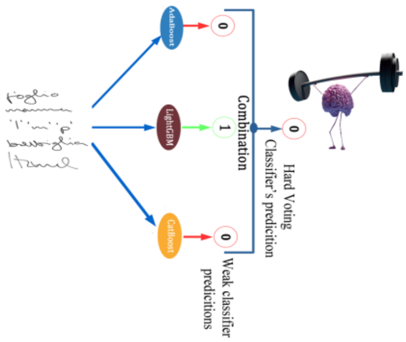

***Figure A**: Architectural structure of the ensemble learning model that detects Alzheimer's disease from handwriting / **Şekil A:**. Alzheimer hastalığını el yazısından tespit eden topluluk öğrenme modelinin mimari yapısı*

Highlights (Önemli noktalar)

- Information about Alzheimer's disease can be obtained based on the deterioration in the patient's writing skills. / Hastanın yazma becerisindeki bozulmaya göre Alzheimer hastalığı hakkında bilgi edinilebilir.
- In this study, Gradient Boosting Machine, Categorical Boosting , and Adaptive Boosting machine learning classification algorithms were combined with a Hard Voting Classifier and trained and tested on the publicly available Diagnosis Alzheimer's With haNdwriting dataset. / Bu çalışmada, Gradient Boosting Machine , Kategorik Boosting ve Adaptive Boosting makine öğrenimi sınıflandırma algoritmaları, Hard Voting Classifier ile birleştirilmiş ve halka açık Diagnosis Alzheimer's With haNdwriting veri kümesi üzerinde eğitilmiştir.

***Aim (Amaç):** The aim of this study is to detect Alzheimer's disease from handwriting quickly and with high sensitivity by combining machine learning-based classifiers. / Bu çalışmanın amacı makine öğrenmesi tabanlı sınıflandırıcıları birleştirerek Alzheimer hastalığını el yazısından hızlı ve yüksek hassasiyet ile tespit etmektir.*

***Originality (Özgünlük):** Employed classification models were used together for the first time in this study. / Kullanılan sınıflandırma modelleri ilk kez bu çalışmada bir arada kullanılmıştır.*

Results (Bulgular): As a result of the experimental studies, the proposed Ensemble methodology achieved 97.14% Acc, 95% Prec, 100% Recall, 90.25% Spec, and 97.44% F1-score (Dice) performance values. */ Deneysel çalışmalar sonucunda önerilen Ensemble metodolojisi %97,14 Acc, %95 Prec, %100 Recall, %90,25 Spec ve %97,44 F1-score (Dice) performans değerlerine ulaştı.*

***Conclusion (Sonuç):** As a result, the proposed methodology showed higher performance than other approaches and individual learning models in the literature. / Sonuç olarak, önerilen metodoloji literatürdeki diğer yaklaşımlar ve bireysel öğrenme modellerine göre daha yüsek performans göstermiştir.*

*Corresponding author, e-mail: hocal@bartin.edu.tr
DOI: 10.29109/gujsc.1386416

GU J Sci, Part C, 12(1): 214-223 (2024)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

A Novel Approach to Detection of Alzheimer's Disease from Handwriting: Triple Ensemble Learning Model

Hakan ÖCAL[1*]

[1]*Bartın University, Faculty of Engineering Architecture and Design, Department of Computer Engineering, Bartın, Turkey*

Article Info

Research article
Received: 05/10/2023
Revision: 12/02/2024
Accepted: 13/02/2024

Keywords

Neurodegenerative disease
Alzheimer's disease prediction
Ensemble machine learning model
Classification
Handwriting dataset

Abstract

The irreversible degeneration of nerve cells in the body dramatically affects the motor skills and cognitive abilities used effectively in daily life. There is no known cure for neurodegenerative diseases such as Alzheimer's. However, in the early diagnosis of such diseases, the progression of the disease can be slowed down with specific rehabilitation techniques and medications. Therefore, early diagnosis of the disease is essential in slowing down the disease and improving patients' quality of life. Neurodegenerative diseases also affect patients' ability to use fine motor skills. Losing fine motor skills causes patients' writing skills to deteriorate gradually. Information about Alzheimer's disease can be obtained based on the deterioration in the patient's writing skills. However, manual detection of Alzheimer's disease (AD) from handwriting is a time-consuming and challenging task that varies from physician to physician. Machine learning-based classifiers are exceptionally popularly used with high-performance scores to solve the difficulty of manual detection of AD. In this study, Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Adaptive Boosting (AdaBoost) machine learning classification algorithms were combined with a Hard Voting Classifier and trained and tested on the publicly available DARWIN (Diagnosis Alzheimer's With haNdwriting) dataset. As a result of the experimental studies, the proposed Ensemble methodology achieved 97.14% Acc, 95% Prec, 100% Recall, 90.25% Spec, and 97.44% F1-score (Dice) performance values. Studies have shown that the proposed research is exceptionally robust.

Alzheimer Hastalığının El Yazısından Tespitinde Yeni Bir Yaklaşım: Üçlü Topluluk Öğrenme Modeli

Makale Bilgisi

Araştırma makalesi
Başvuru: 05/10/2023
Düzeltme: 12/02/2024
Kabul: 13/02/2024

Anahtar Kelimeler

Nörodejeneratif hastalık
Alzheimer hastalığı tahmini
Topluluk makine öğrenimi modeli
Sınıflandırma
El yazısı veri kümeleri

Öz

Vücuttaki sinir hücrelerinin geri dönüşü olmayan dejenerasyonu, günlük yaşamda etkin olarak kullanılan motor becerileri ve bilişsel yetenekleri dramatik biçimde etkiler. Alzheimer gibi nörodejeneratif hastalıkların bilinen bir tedavisi yoktur. Ancak bu tür hastalıkların erken teşhisinde spesifik rehabilitasyon teknikleri ve ilaçlarla hastalığın ilerlemesi yavaşlatılabilir. Bu nedenle hastalığın erken tanısı, hastalığın yavaşlatılması ve hastaların yaşam kalitesinin artırılması açısından önemlidir. Nörodejeneratif hastalıklar aynı zamanda hastaların ince motor becerilerini kullanma yeteneğini de etkiler. İnce motor becerilerin kaybı hastaların yazma becerilerinin giderek bozulmasına neden olur. Hastanın yazma becerisindeki bozulmaya göre Alzheimer hastalığı hakkında bilgi edinilebilir. Ancak Alzheimer hastalığının (AH) el yazısından manuel olarak tespiti, hekimden hekime değişen, zaman alıcı ve zorlu bir iştir. Makine öğrenimi tabanlı sınıflandırıcılar, AD'nin zor manuel tespitini çözmek için yüksek performanslı puanlarla son derece popüler bir şekilde kullanılır. Bu çalışmada, Light Gradient Boosting Machine (LightGBM), Kategorik Boosting (CatBoost) ve Adaptive Boosting (AdaBoost) makine öğrenimi sınıflandırma algoritmaları, Hard Voting Classifier ile birleştirildi ve halka açık Diagnosis Alzheimer's With haNdwriting (DARWIN) veri kümesi üzerinde eğitildi ve test edildi. . Deneysel çalışmalar sonucunda önerilen Ensemble metodolojisi %97,14 Acc, %95 Prec, %100 Recall, %90,25 Spec ve %97,44 F1-score (Dice) performans değerlerine ulaştı. Çalışmalar, önerilen çalışmanın son derece sağlam olduğunu göstermiştir.

1. INTRODUCTION (GİRİŞ)

The human brain is an organ that fulfills vital functions such as thinking, decision-making, and storage of experiences in their memory [1]. Procedures are performed through nerve cells in the

*Corresponding author, e-mail: hocal@bartin.edu.tr
DOI: 10.29109/gujsc.1386416

human brain. Neurodegenerative conditions in nerve cells are irreversible circumstances that gradually deteriorate the individual's quality of life. The leading neurodegenerative disease is AD, which causes the death of memory cells and gradual shrinkage of the brain [2].

In 2019, in the United States, AD caused the death of 121,499 people. In the year, COVID-19 ranks 10th among deadly diseases in the USA, while AD ranks 6th. Improving the quality of life of individuals will increase life expectancy, and the number of individuals with AD will gradually increase in the coming years. Increasing the number of individuals with AD will lead to insufficient magnetic resonance (MR) and other costly diagnostic techniques. The inadequacy of the AD diagnostic procedures used today due to financial opportunities leads scientists to research new and less expensive diagnostic methods. Scientists have explicitly focused on the fact that AD causes losses in the individual's fine motor skills. Therefore, they thought diagnosing the severity of AD by monitoring the deterioration in an individual's handwriting could be a noninvasive method that does not require external intervention for the patient [3,4].

Machine learning applications have become extremely popular in recent years due to their high performance in detecting diseases. Manual AD detection from handwriting is exceptionally time-consuming and challenging for physicians. This study focused on a high-performance machine learning model for diagnosing Alzheimer's disease using handwriting to solve the challenges of manual AD detection using patient handwriting. As mentioned in the methodology section of the article, experimental studies have also shown that if the models that make up an ensemble learning-based machine learning model are fine-tuned effectively, they will be compatible with each other, and a higher-performance machine learning model will be obtained.

To briefly summarize the contribution of this study to the literature:

- In this study, Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Adaptive Boosting (AdaBoost) powerful machine learning models, which have high-performance results in the literature, were combined for the first time to detect AD from handwriting.

- Also, 10-fold cross-validation was performed when selecting machine learning models that make up the ensemble model.

- In addition, by fine-tuning, the individual performances of the models were brought closer together, enabling them to make more errors and achieve a higher performance score.

Additionally, to calculate the inter-case variance of handwriting tasks, the extracted features for each task were subjected to Principal Component Analysis (PCA) [9].

In the 2nd part of the study, summary information about the relevant studies will be given. In contrast, in the 3rd part, information about the data style and machine learning methods used will be provided. In Section 4, the analysis results of machine learning algorithms on the DARWIN dataset will be compared and discussed. In the 5th section, the last part of the study, the research results, and future studies will be shared.

In recent years, machine learning methods have been increasingly used to solve many problems [10-12]. Machine learning models have also achieved many successes in the field of healthcare. They have become popular in healthcare, especially as computer-aided systems in diagnosing neurodegenerative diseases (ND). The most important and necessary thing for machine learning models is the quality and size of the dataset. For the detection of Parkinson's disease (PD) from handwriting, there are robust public datasets such as the Parkinson's Disease Handwriting Database (PaHaW) and HandPD [13, 8].

There are various studies in the literature on the detection of PD disease from handwriting [[14-19]. However, the fact that the data sets obtained on a case-by-case basis are only for the detection of PD disease has caused insufficient studies on the detection of AD from handwriting. However, there are few datasets in the literature for detecting AD from handwriting. Therefore, the DARWIN handwriting dataset, consisting of 174 participants and based on 25 handwriting tasks, was created by Cilia et al. The resulting dataset was benchmarked against nine different machine-learning models. These models were tested separately for 25 various tasks, and an Ensemble model called BFT was created, combining the results [20].

One of the limited studies in the literature on detecting Alzheimer's disease was conducted by Chai et al. In this study, 75 handwritten and quantitative electroencephalography (qEEG) data obtained from 30 healthy individuals and 40 individuals with mild cognitive impairment (MCI) were tested in the Support Vector Machine (SVM) machine learning algorithm with RBF kernels [21]. In another study, El-Yacoubi et al. used two-stage semi-supervised learning and word-based feature extraction methods to detect early-stage Alzheimer's disease (ES-AD) and early-stage Alzheimer's disease (ES-AD) in the Ironoff dataset consisting of 880 handwritings and 25 undiagnosed cases of AD obtained from Broca Hospital. They tried to determine MCI. In the Ironoff dataset, the number of cases over 60 is meager. Therefore, to make the study results more reliable, 25 patients with an average age of 72 without a diagnosis of AD were obtained from Broca Hospital [22]. Kahindo et al. tried to predict the classes of the subjects by using Hierarchical Clustering and K-means algorithms for feature selection on 144 subjects, including ES-AD, MCI, and HC individuals obtained from Broca Hospital. Additionally, the Normalized Mutual Information (NMI) metric and semi-global feature parametrization technique were used in this study [23]. For detecting PD from handwriting, Sarin et al. proposed a fuzzy classification method consisting of 3 stages for detecting Parkinson's disease from handwriting [24].

The proposed study, which can be seen from the literature, is considered one of the pioneer studies. The reason for this is the abundance of publicly available handwriting datasets for detecting PD disease and the inadequacy of AD datasets so far. The handwriting dataset for detecting high-incidence AD, introduced to the literature by Cilia et al., constitutes a cornerstone for studies to be carried out for less costly detection of AD.

2. MATERIALS AND METHODS (MATERYAL VE METOD)

This section provides information about the dataset and machine learning methodologies used.

2.1. Preparing the dataset (Veri setinin ön hazırlığı)

The proposed study used the DARWIN dataset for comparative analysis of machine learning models. The dataset is the largest publicly available dataset used for detecting AD, with 25 different tasks and 174 participants. Of the participants in the dataset used, 89 were AD patients, and 85 were healthy individuals. For the training of the ensemble model, the DARWIN data set was randomly divided into 80% training and 20% test data using the model selection method of the Sckit-learn library. The most significant impact of the dataset used is that it eliminates the scarcity of data in MR images, which is another method used to diagnose AD. Handwritten data in the DARWIN dataset were performed according to the acquisition protocol proposed by Cilia et al. [20]. 25 tasks in the dataset are grouped into three categories.

• Graphic tasks (G) consist of the participant creating geometric shapes by connecting dots and labeling these shapes with basic writing skills.

•Copying tasks (C) consist of the participant's ability to repeat semantic symbols such as letters, words, and numbers.

• It consists of memory and dictation tasks(M) that question the differences in the writing process that have previously been memorized or associated objects in a picture and how the handwriting in working memory changes.

Standard clinical tests such as the Mini-Mental State Examination (MMSE), Preliminary Assessment Battery (FAB), and Montreal Cognitive Assessment (MoCA) were used to recruit participants who comprised the dataset. These tests used questionnaires covering many cognitive skills, including time and space orientation remembering skills. Gender, age, education, and job levels are equally distributed in the dataset. A total of 25 tasks in groups C, G, and M of the data set used in the proposed study are shown in Figure 1. It can be seen that the 25 features used consist of various writing and drawing tasks. As can be seen from the figure, 14 tasks for group C, 6 tasks for group G and 5 tasks for group M were determined in the data set.

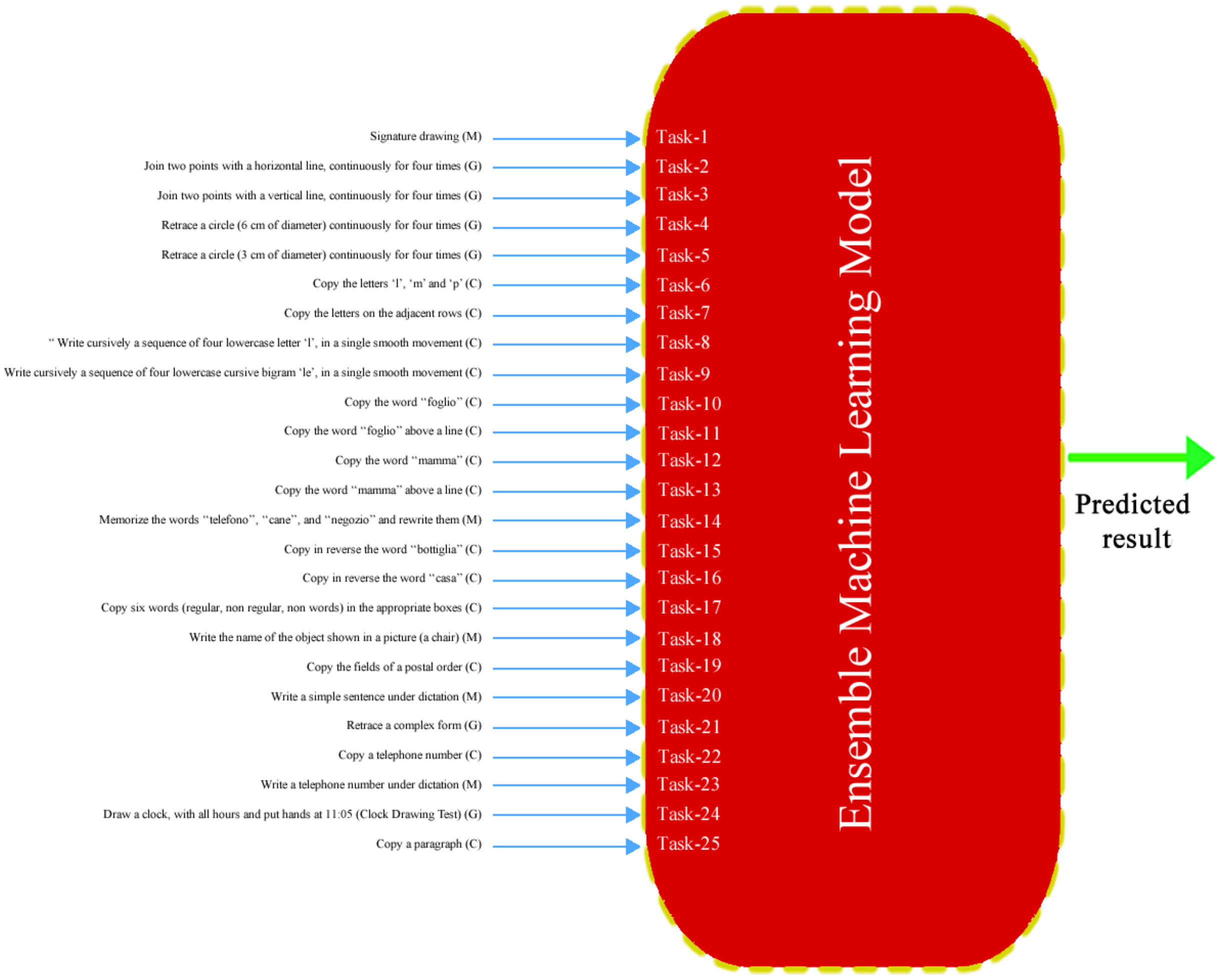

Figure 1. Block diagram of the operation of a 25-task classifier (25 görevli bir sınıflandırıcının çalışmasının blok diyagramı)

2.2. Employed machine learning methodologies
(Kullanılan makine öğrenimi metodolojileri)

The most commonly used machine learning-based classification models in the literature, Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Adaptive Boosting (AdaBoost), were combined to classify 25 different tasks in the data set used. As shown in Section 4, various ablation studies have been performed on recruited independent classification models. The models that form the ensemble model used to classify the data set achieved higher performance than other machine learning algorithms in solving different problems. Individual classification models were implemented in Python using the Scikit-Learn, CatBoost, and Lightgbm libraries. Table 1 shows the computational complexities of the selected classification models. LightGBM is a histogram-based classification model developed by Microsoft in 2017. LightGBM is an algorithm designed to deal with big data. The algorithm makes continuous data discrete by dividing it into nodes. In this way, it dramatically reduces the data size and the number of features, significantly reducing training time and parameter usage. LightGBM has been found to be 20 times faster than other classification algorithms in studies (A Highly Efficient Gradient Boosting Decision Tree). Since the LightGBM model has a leaf-oriented learning strategy, it makes fewer errors and learns faster. However, the leaf-oriented learning strategy is more prone to overlearning when data is scarce. Therefore, overlearning can be prevented in low data by optimizing parameters such as learning rate, tree depth, and number of leaves. Figure 2 shows LightGBM's classification strategy.

For data classification, LightGBM is Gradient-based One-Way Sampling (GOSS), which focuses on data samples, and Exclusive Feature Bundling (EFB), which deals with the number of variables.

GOSS is a method that preserves the accuracy of decision trees and cleans unwanted data from the data by looking at their gradients, thus reducing the number of data. GOSS ensures that the machine learning algorithm focuses only on high-value features. EFB, on the other hand, combines sparse features using a leaf-wise growth strategy to reduce dimensionality. Accordingly, complexity is reduced, and training time is shortened.

Table 1. Abbreviations and time complexity (O notation) of the training phase of the classification models used. *N* represent the number of training samples. As for the other quantities involved, they are described as follows: (Kullanılan sınıflandırma modellerinin eğitim aşamasının kısaltmaları ve zaman karmaşıklığı (O notasyonu). *N* eğitim örneklerinin sayısını temsil eder. İlgili diğer miktarlara gelince, bunlar aşağıdaki gibi tanımlanmaktadır:)

T: Number of weak learners;

f: weak learner in use;

TS: Target Statistics

Methodologies	Abbreviations	Time Complexity
Light Gradient Boosting Machine	LGBM	0.5 * #feature * #bin
Adaptive Boosting	AdaBoost	T f
Categorical Boosting	CatBoost	$N_{TS,t} \cdot n$

The AdaBoost (short for Adaptive Boosting) classification algorithm is a popular machine learning algorithm introduced by Yoav Freund and Robert Schapire in 1995. The AdaBoost machine learning model combines the outputs of weak classifiers to build a robust classification model. Weak classifiers try to minimize the misclassification rate of previous weak classifiers on the training data. For this, the AdaBoost algorithm re-weights the dataset before each weak classifier and feeds it to the weak classifier, as seen in Figure 3. Iteration and rounding of these weights continue according to the determined number of weak classifiers. The values obtained from the weak classifier are fed to a non-linear Ensemble classifier. According to the training result obtained from the ensemble classifier, the error is reduced by increasing the weights of the incorrectly predicted training samples. The weight value of each weak classifier is increased or decreased according to its accuracy rate. The weight value of a weak classifier with a high accuracy rate is also high. The model's tendency to overfit is also relatively low. The AdaBoost algorithm can be pretty sensitive to noisy samples and outliers. However, it is quite successful in analyzing large and complex data.

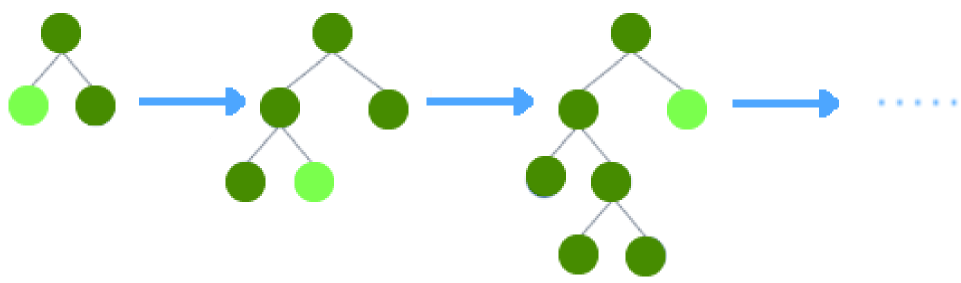

Figure 2. LightGBM Leaf-wise tree growth (LightGBM Yaprak şeklinde ağaç büyümesi)

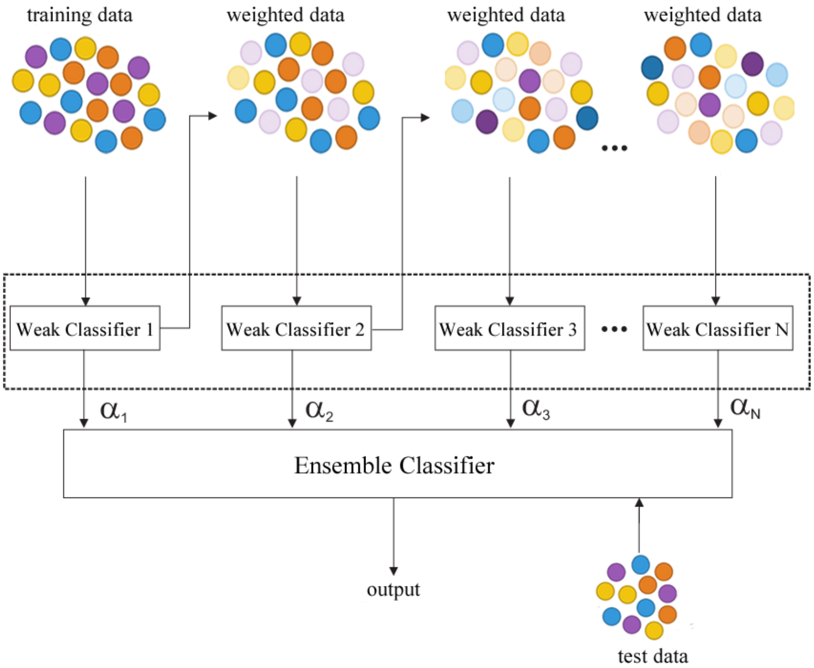

Figure 3. AdaBoost block diagram (AdaBoost blok diyagramı)

The CatBoost machine learning algorithm, seen in Figure 4, is effective in regression, classification, and multidimensional classification. The gradient gradient of these methods may differ depending on the objective function. Additionally, the Catboost algorithm has built-in a priori metrics to obtain the best testing performance before performing performance evaluation on the data set. The CatBoost algorithm reduces the error by creating several binary decision trees simultaneously. As its name suggests, it is an algorithm that performs highly on categorical data. In addition, the CatBoost algorithm performs more in dealing with overfitting in small data sets due to its data pre-processing feature. Using the one_hot_max_size method, Catboost retrieves all features with many distinct values less than or equal to the feature parameter value given to the model. Thus, it obtains high-level features more quickly. Additionally, CatBoost is grouped by target statistics (TS), estimating each category's expected target value. In CatBoost, the data is constantly mixed throughout the training, and the average value is calculated for each category.

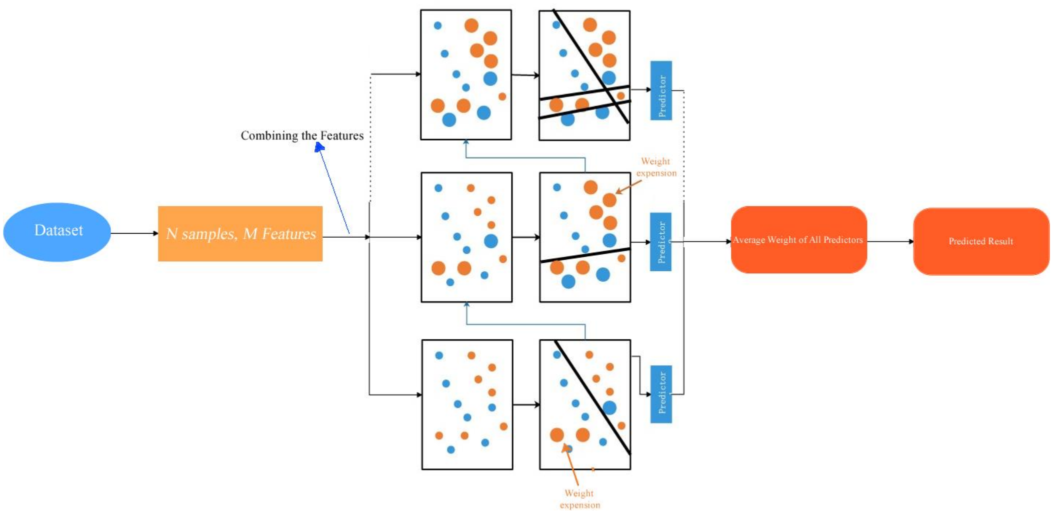

Figure 4. CatBoost block diagram (CatBoost blok diyagramı)

2.3. Proposed methodology (Önerilen metodoloji)

The proposed model is a binary classification model that distinguishes Alzheimer's patients from healthy people. In the proposed methodology, AdaBoost, LightGBM, and CatBoost classification algorithms are combined with the help of a Hard Voting classifier. The proposed method is shown in Figure 5. A hard voting classifier is a machine learning model that predicts the outcome based on the highest probability of the selected class from the classifiers in a machine learning-based ensemble group of many models. The classification model based on the majority of the predictions coming from the classifiers is called hard voting classification. As can be seen in the experimental studies of the proposed model, the hard voting classifier model based on hard voting classification can achieve higher classification results than the classifiers in the community group.

In this model, where the models forming the ensemble are weak learners, much higher prediction scores can be obtained if there are sufficient weak learners (3 or more classifiers, according to our study).

In the hard voting classifier model, classifier models that are as independent from each other as possible should be selected in selecting the classifier that forms the model. Independent machine learning algorithms increase the error rate of classifiers and reduce overfitting.

In the hard voting classifier model, classifier models that are as independent from each other as possible should be selected in selecting the classifier that forms the model. Independent machine learning algorithms increase the error rate of classifiers and reduce overfitting.

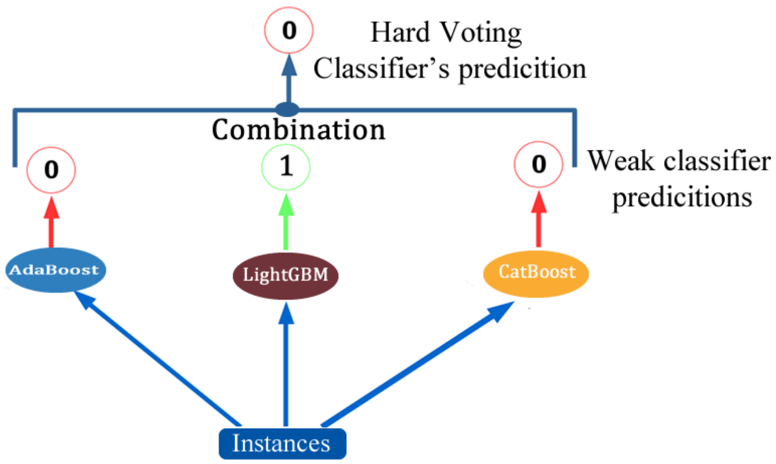

Figure 5. Block diagram of the proposed methodology (Önerilen metodolojinin blok diyagramı)

3. RESULTS (BULGULAR)

3.1. Performance metrics (Performans metrikleri)

The performance evaluation of the proposed ensemble classifier was carried out using Accuracy(Acc), Precision(Prec), Recall, Specificity(Spec), and F1-score(F1) metrics. Performance metrics provide insight into the quantitative limitations of an architecture. The mathematical models of the proposed performance metrics are shown in equations 1, 2, 3, 4, and 5. The True Positive (TP) value in the equations shows the test examples where the prediction result of the model is positive, and the sample in the class is positive. True Negative (TN) is when the true value of the test sample is negative, and the predicted result is also negative. False Positive (FP) is when the actual test value is negative, and the predicted test sample result is positive. False Negative refers to situations where the ground truth is positive and the predicted result is negative.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall} \quad (5)$$

3.2. Ablation study (Ablasyon çalışmaları)

CatBoost, AdaBoost, and LightGBM machine learning models achieved higher performances than other machine learning models on the DARWIN dataset used in the 10-fold cross-validation tests. The success of these models in previous studies on data sets with many features is another point that encourages using these three models together in the study. In addition, when more than three classifiers was tried, the performance of the ensemble model decreased. Also, various ablation studies have been conducted to obtain the best classification results in the classification algorithms that form the proposed ensemble learning-based hard voting classifier model. In the LightGBM classifier model, hyperparameters other than the learning rate did not contribute to increasing sensitivity. When the learning rate was selected as 0.5, 91.43% Accuracy, 90% Precision, 94.74% Recall, and 92.31% F1-score values were obtained. When the number of estimators(n_estimators)=100 and learning rate 100 were selected in the optimization studies of the AdaBoost classification algorithm on the dataset, 94.29% Accuracy, 94.74% Precision, 94.74% Recall, and 94.74% F1-score performance results were obtained. A better performance value could not be obtained than the performance values obtained in the studies of increasing and decreasing the learning rate and n_estimators values. Hyperparameters other than n_estimators and learning rate did not have any effect on improving the performance of the AdaBoost classification algorithm. In the performance studies of the CatBoost classification algorithm on the dataset, 97.14% Accuracy, 100% Precision, 94.74% Recall, and 94.30% F1-score values were obtained when learning rate=0.5 and depth=5 were selected. However, the community classification result obtained from the hard voting classification model was 94.29% Accuracy, 94.74% Precision, 94.74% Recall, and 94.74% F1-score. The high-performance result achieved by CatBoost prevented the ensemble model from making more errors and reduced its learning performance. When learning rate=0.6 and depth=6 were selected in the CatBoost classifier, CatBoost alone performance achieved 91.43% Accuracy, 86.36% Precision, 1.0% Recall and 92.68% F1-score values. However, the decrease in CatBoost's performance enabled the ensemble model to achieve 97.14% Accuracy, 95% Precision, 100% Recall, and 97.44% F1-score values.

3.3. Comparative analysis of the proposed model with state-of-the-art approaches (Önerilen modelin en son teknoloji yaklaşımlarla karşılaştırmalı analizi)

A comparative analysis of the proposed ensemble model with single models and other studies in the literature is shown in Table 2. As can be seen in Table 2, the proposed methodology achieved superior success compared to other machine learning models on the DARWIN dataset. As a result of ablation studies, the ensemble model obtained as a result of fine adjustments of the LightGBM, AdaBoost, and CatBoost models achieved a performance score of 6 points higher in Acc, 0.5 points in Prec, and 2 points in Spec than the ensemble classification architecture consisting of 9 classifiers proposed by Cilia et al.

Hard voting outputs predictions based on a majority vote from the predictions of independent classifiers. For high performance in hard voting classifiers, predictive classification algorithms should be as independent and different from each other as possible. Therefore, it was adopted as the primary algorithm choice, and the classification algorithms used in this study were independent of each other. Suppose the individual performances of the models are brought closer to each other by fine-grain tuning. In that case, the error between the independent algorithms will increase, and the accuracy of training and testing will be higher.

If a good fit is achieved in the independent algorithms, the hard voting classification technique will achieve success superior to the individual success of the independent models. In addition, since the data set used in the proposed study requires a simpler model, hard voting achieved higher performance than soft voting. Adding more models to the ensemble also reduced learning, resulting in unsuccessful test results. Different machine learning models, such as SVM, Random forest, and Decision Tree, have also been added to the machine learning models in the proposed architecture. However, the ensemble learning model has achieved poor test performance.

Table 2. Performance analysis of the proposed ensemble model compared to other methodologies
(Önerilen topluluk modelinin diğer metodolojilerle karşılaştırıldığında performans analizi)

Dataset	Classifier	Acc (%)	Prec (%)	Recall (%)	Spec (%)	F1 (%)
DARWIN	AdaBoost	94.29	94.74	94.74	89.64	94.74
	LightGBM	91.43	90	94.74	88.27	92.31
	CatBoost	91.43	86.36	1.0	81.45	92.68
	BFT [14]	91.43	94.44	-	88.24	-
	Ensemble(Proposed)	**97.14**	**95**	**1.0**	**90.25**	**97.44**

4. CONCLUSIONS (SONUÇLAR)

This study proposes an ensemble learning model combining powerful machine learning-based classification algorithms such as AdaBoost, CatBoost, and LightGBM to detect Alzheimer's disease from handwriting. The most important feature of this study is that it is a pioneering study in the literature for detecting AD from handwriting. The low number of cases in the data sets before the DARWIN data set did not allow the detection of AD from handwriting. The publicly available DARWIN dataset was used to train and test the proposed methodology. In ensemble learning, the Hard Voting Classifier classification algorithm was employed to produce a result based on the predictions from weak classifiers. Various ablation studies were carried out individually on weak classifier models to obtain the most robust and high-performance ensemble model. Experimental studies were scored comparatively with multiple performance metrics. As a result of the experimental studies, the proposed Ensemble methodology achieved 97.14% Acc, 95% Prec, 100% Recall, 90.25% Spec, and 97.44% F1-score (Dice) performance values. Studies have shown that the proposed work is exceptionally robust.

DECLARATION OF ETHICAL STANDARDS (ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

***Hakan ÖCAL*:** He conducted the experiments, analyzed the results and performed the writing process.

Deneyleri yapmış, sonuçlarını analiz etmiş ve maklenin yazım işlemini gerçekleştirmiştir.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] Armstrong MJ, Litvan I, Lang AE, Bak TH, Bhatia KP, Borroni B, Boxer AL, Dickson DW, Grossman M, Hallett M, Josephs KA, Kertesz A, Lee SE, Miller BL, Reich SG, Riley DE, Tolosa E, Tröster AI, Vidailhet M, Weiner WJ. Criteria for the diagnosis of corticobasal degeneration. Neurology. 2013; 80 (5), 496–503.

[2] Alberdi A, Weakley A, Schmitter-Edgecombe M, Cook DJ, Aztiria A, Basarab A, Barrenechea M. Smart home-based prediction of multidomain symptoms related to Alzheimer's disease, IEEE J. Biomed. Health Inform. 2018; 22:1720–31.

[3] Precup RE, Teban TA, Albu A., Borlea AB, Zamfirache IA, Petriu EM. Evolving fuzzy models for prosthetic hand myoelectric-based control. IEEE Trans. Instrum. Meas. 2020; 69: 4625–4636.

[4] Impedovo D, Pirlo G, Vessio G. Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. Information. 2018; 9(10), 247.

[5] Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, Hautbergue GM, Holbrook JD, Ferraiuolo L. Applications of machine learning to

diagnosis and treatment of neurodegenerative diseases. Nat. Rev. Neurol. 2020; 16: 440–456.
[6] Albu A, Precup R, Teban T. Results and challenges of artificial neural networks used for decision making and control in medical applications. Facta Univ. Ser.: Mech. Eng. 2019; 17(3): 285-308.
[7] Tanveer M, Richhariya B, Khan RU, Rashid AH, Khanna P, Prasad M, Lin CT. Machine learning techniques for the diagnosis of Alzheimer's disease: A review. ACM Trans. Multimedia Comput. Commun. Appl. 2020; 16(1): 1-35.
[8] Pereira CR, Pereira DR, Weber SA, Hook C, de Albuquerque VHC, Papa J. A survey on computer-assisted Parkinson's disease diagnosis, Artif. Intell. Med. 2019; 95: 48–63.
[9] Ghaderyan P, Abbasi A, Ebrahimi A, Time-varying singular value decomposition analysis of electrodermal activity: A novel method of cognitive load estimation, Measurement, 2018; 126: 102-109.
[10] Jain N, Virmani D, Abraham A. Proficient 3-class classification model for confident overlap value-based fuzzified aquatic information extracted tsunami prediction. Intell. Decis. Technol. 2019; 13: 295–303.
[11] Borlea ID, Precup RE, Borlea AB, Iercan D. A unified form of fuzzy C-means an K-means algorithms and its partitional implementation. Knowl.-Based Syst. 2021; 214: 106731.
[12] Tonkal Ö, Polat H. Traffic Classification and Comparative Analysis with Machine Learning Algorithms in Software Defined Networks. Gazi University Journal of Science Part C: Design and Technology. 2021; 9(1): 71-83.
[13] Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. Artif. Intell. Med. 2016; 67: 39–46.
[14] Loconsole C, Cascarano GD, Brunetti A, Trotta GF, Losavio G, Bevilacqua V, Di Sciascio E. A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis. Pattern Recognition Letters. (2019); 121: 28-36.
[15] Moetesum M, Siddiqi I, Vincent N, Cloppet F. Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease. Pattern Recognition Letters. 2018; 121: 19-27.
[16] Slavin MJ, Phillips JG, Bradshaw JL, Hall KA, Presnell I. Consistency of handwriting movements in dementia of the Alzheimer's type: a comparison with Huntington's and Parkinson's diseases. J. Int. Neuropsychol. 1999; 5(1): 20–25.
[17] Teulings HL, Stelmach GE. Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting. Hum. Mov. Sci. 1991; 10 (2): 315–334.
[18] Thomas M, Lenka A, Kumar Pal P. Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions. Movement Disorders Clinical Practice. 2017; 4(6): 806-818.
[19] Walton J, Handwriting changes due to aging and Parkinson's syndrome. Forensic Sci. Int. 1997; 88(3): 197-214.
[20] Cilia ND, De Gregorio G, De Stefano C, Fontanella F, Marcelli A, Parziale A, Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. Engineering Applications of Artificial Intelligence. 2022; 111: 0952-1976.
[21] Chai J, Wu R, Li A, Xue C, Qiang Y, Zhao J, Zhao Q, Yang Q. Classification of mild cognitive impairment based on handwriting dynamics and qEEG. Comput Biol Med. 2023; 152:106418.
[22] El-Yacoubi MA, Garcia-Salicetti S, Kahindo C, Rigaud AS, Cristancho-Lacroix V. From aging to early-stage Alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning, Pattern Recognition. 2019; 86: 112-133.
[23] Kahindo C, El-Yacoubi MA, Garcia-Salicetti S, Cristancho-Lacroix V, Kerhervé H, Rigaud AS. Semi-global Parameterization of Online Handwriting Features for Characterizing Early-Stage Alzheimer and Mild Cognitive Impairment. IRBM. 2018; 39(6): 421-429.
[24] Sarin K, Bardamova M, Svetlakov M, Koryshev N, Ostapenko R, Hodashinskaya A, Hodashinsky I. A three-stage fuzzy classifier method for Parkinson's disease diagnosis using dynamic handwriting analysis. Decision Analytics Journal. 2023; 8: 100274.