

Article

Noninvasive Deep Learning Analysis for Smith–Magenis Syndrome Classification

Esther Núñez-Vidal ^{1,*} , Raúl Fernández-Ruiz ¹ , Agustín Álvarez-Marquina ² , Irene Hidalgo-delaGuía ³ , Elena Garayzábal-Heinze ⁴ , Nikola Hristov-Kalamov ¹ , Francisco Domínguez-Mateos ¹ , Cristina Conde ¹  and Rafael Martínez-Olalla ² 

¹ Escuela Técnica Superior de Ingeniería Informática, Universidad Rey Juan Carlos, C/Tulipán S/N, 28933 Móstoles, Spain; raul.fruiz@urjc.es (R.F.-R.); nikola.hristov@urjc.es (N.H.-K.); francisco.dominguez@urjc.es (F.D.-M.); cristina.conde@urjc.es (C.C.)

² Center for Biomedical Technology, Universidad Politécnica de Madrid, 28220 Madrid, Spain; aalvarez@fi.upm.es (A.Á.-M.); rmolalla@fi.upm.es (R.M.-O.)

³ Departament of Spanish Language and Theory of Literature, Universidad Complutense de Madrid, 28040 Madrid, Spain; irhidalg@ucm.es

⁴ Department of Linguistics, Universidad Autónoma de Madrid, 28049 Madrid, Spain; elena.garayzabal@uam.es

* Correspondence: e.nunezv.2021@alumnos.urjc.es

Abstract: Smith–Magenis syndrome (SMS) is a rare, underdiagnosed condition due to limited public awareness of genetic testing and a lengthy diagnostic process. Voice analysis can be a noninvasive tool for monitoring and detecting SMS. In this paper, the cepstral peak prominence and mel-frequency cepstral coefficients are used as disease monitoring and detection metrics. In addition, an efficient neural network, incorporating synthetic data processes, was used to detect SMS in a cohort of individuals with the disease. Three study cases were conducted with a set of 19 SMS patients and 292 controls. The three study cases employed various oversampling and undersampling techniques, including SMOTE, random oversampling, NearMiss, random undersampling, and 16 additional methods, resulting in balanced accuracies ranging from 69% to 92%. This is the first study using a neural network model to focus on a rare genetic syndrome using phonation analysis data. By using synthetic data (oversampling and undersampling) and a CNN, it was possible to detect SMS with high levels of accuracy. Voice analysis and deep learning techniques have proven to be a useful and noninvasive method. This is a finding that may help in the complex identification of this syndrome as well as other rare diseases.

Keywords: CNN; deep learning; Smith–Magenis syndrome; speech; synthetic data



Citation: Núñez-Vidal, E.; Fernández-Ruiz, R.; Álvarez-Marquina, A.; Hidalgo-delaGuía, I.; Garayzábal-Heinze, E.; Hristov-Kalamov, N.; Domínguez-Mateos, F.; Conde, C.; Martínez-Olalla, R. Noninvasive Deep Learning Analysis for Smith–Magenis Syndrome Classification. *Appl. Sci.* **2024**, *14*, 9747. <https://doi.org/10.3390/app14219747>

Academic Editor: Christos Bouras

Received: 15 September 2024

Revised: 1 October 2024

Accepted: 8 October 2024

Published: 25 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human voice carries a wealth of information beyond just the spoken word. It can reveal characteristics related to a speaker’s anatomy, age, gender, and even lifestyle. But beyond these speaker-specific traits, voice analysis, particularly the study of how sound is produced (phonation), offers valuable insights into an individual’s health. Neurological pathologies can be identified by analyzing phonation patterns [1,2]. Disruptions in the brain’s control of speech production mechanisms, caused by neurological disorders, manifest as changes in the voice source (glottis). This situation results in dysarthria, which is a speech disorder characterized by difficulty forming clear and coordinated speech. Some of the most common neurological causes of dysarthria include stroke, traumatic brain injury (TBI), brain tumors, multiple sclerosis, Alzheimer’s dementia, Parkinson’s disease, amyotrophic lateral sclerosis (ALS), or effects resulting from rare genetic disorders that cause a variety of developmental and physical problems, among others. According to neurological and neurodegenerative diagnosis and monitoring have witnessed a recent surge in utilizing artificial intelligence (AI), particularly with the rise of deep learning

techniques [3]. This powerful approach focuses on specific conditions like Parkinson's, Huntington's, Alzheimer's, and schizophrenia. By analyzing data such as facial expressions, speech patterns [4], movement [5,6], brain scans (MRI) [7], cognitive impairments [8,9], mental disorders [10], and motor symptoms [11,12], deep learning models offer valuable insights to aid in diagnosing these diseases [13,14]. Likewise, voice data analysis using deep learning models presents a promising approach for early detection and disease progression assessment. This method offers several advantages, including cost-effectiveness and reduced invasiveness, compared to traditional methods. While a substantial body of research exists on deep learning and phonation analysis, particularly in the context of Parkinson's disease [15–18], its application to other pathologies remains largely unexplored. This fact presents a significant opportunity to address the underdiagnosis of numerous genetically linked syndromes.

Scientific work on genetic diseases has focused on analyzing medical data, genetic tests, and diagnostic procedures, often invasive. Rare diseases affect a small portion of the population and collectively pose a significant challenge to global healthcare. In the United States, a disease is rare if it affects fewer than 200,000 individuals, with an estimated 30 million people living with a rare disease according to the National Organization for Rare Disorders (NORD) [19]. Limited patient numbers often result in insufficient funding for rare disease research, hindering treatment development.

In Europe, rare diseases present a unique healthcare challenge, with up to 36 million people in the EU affected, as reported by the European Commission [20]. In Spain, over 3 million individuals have rare diseases, representing around 7% of the population. The project Spain-RDR (Spanish Rare Diseases Registries Research Network) aims to build the National Registry of Rare Diseases in Spain, coordinated by the Institute of Rare Diseases Research and involving collaboration from various healthcare departments, patient organizations, medical societies, research networks, and the pharmaceutical industry [21]. Spain-RDR focuses on patient-oriented research outcome registries and population-based registries for epidemiological and socio-health research and health planning.

Organizations like NORD, Rare Diseases International (RDI), European Organisation for Rare Diseases (EURORDIS) [22], or Spanish Federation of Rare Diseases (FEDER) advocate for increased research funding and support for rare disease patients. Rare diseases often have a profound impact on quality of life, requiring ongoing medical care and management.

Early diagnosis is vital for less common genetic and neurodevelopmental syndromes like Williams, Prader–Willi, Fragile X, Smith–Magenis, and Angelman syndromes, among many others. Delayed diagnosis poses a significant challenge, resulting in frustrating and prolonged (mis)diagnosis/accurate diagnosis for families. This delay can lead to emotional distress and hinder access to appropriate treatments, which are crucial for improving a patient's quality of life.

Conventional diagnostic methods for genetic syndromes face limitations, driving the search for faster, accessible, and cost-effective alternatives. Voice analysis using advanced deep learning methods emerges as a promising solution. While initially developed for neurodegenerative disorders with severe dysarthria, these methods show potential for detecting subtle phonation deviations in genetic syndromes with milder dysarthria [23–25]. By leveraging deep learning's capacity to analyze complex datasets, these models can identify and utilize fined phonatory differences [26,27]. Such advancements promise earlier and more accurate genetic syndrome diagnoses, potentially transforming the lives of patients and their families [28–30].

Neurodevelopmental delays resulting from chromosomal abnormalities are a primary cause of distinct phonation patterns observed in syndromes. Variations among syndromes are further influenced by additional clinical factors affecting laryngeal biomechanics. Examples include elastin deficiency in Williams syndrome and heightened laryngeal tension in Smith–Magenis syndrome [24,31].

This research work explores the potential of voice analysis, specifically phonation patterns, as a classification tool for genetic syndromes. The core idea lies in the possibility that distinct genetic syndromes manifest with unique phonatory characteristics, similar to how Parkinson's disease is diagnosed through voice analysis [32]. Here, deep learning emerges as a powerful ally. By training neural networks on datasets of voice samples from patients with a genetic syndrome, a network could learn to identify the specific phonatory signatures of the disease. Ideally, this possibility could lead to classifying a particular syndrome based on a single voice recording. Likewise, this study represents a pioneering step towards this ambitious goal.

There are some gaps where the research can contribute valuable insights into the application of noninvasive deep learning techniques for analyzing voice characteristics in individuals with SMS, to enhance the understanding and the development of different treatment techniques. Some of these are the following:

1. The limited research on SMS voice characteristics. This gap highlights the need for more comprehensive research to understand the peculiarities of the patients [30].
2. The lack of noninvasive deep learning applications using phonation data. Research on applying these techniques to SMS voice features is lacking, offering a chance to explore noninvasive deep learning for vocal data analysis [33].
3. Integration of deep learning with clinical practice. There is a gap in using deep learning models in clinical practice for SMS identification or classification. Creating noninvasive deep learning tools to help clinicians monitor SMS voice symptoms could greatly improve patient outcomes.

The objective of this study is to develop a convolutional neural network (CNN)-based classification model capable of accurately identifying individuals with SMS from voice samples. By applying data augmentation techniques, the authors aim to enhance the performance of these models and contribute to the early diagnosis of this disease. Consequently, it is hypothesized that the extraction of discriminative acoustic features such as MFCC and CPP, combined with data augmentation techniques, will enable CNNs to identify underlying patterns in voice signals that differentiate individuals with SMS from healthy controls. In addition, existing voice databases identified in the literature lack data from Spanish-speaking SMS patients. This study aims to fill this gap by using a database that encompasses 20% of SMS patients in Spain. Furthermore, to address the challenge of imbalanced datasets, this research work incorporates sixteen distinct variations of the SMOTE. Each SMOTE variation is employed to generate synthetic data points for the minority class, effectively balancing the class distribution within the dataset.

The paper is organized as follows: the state of the art is presented in the following section. Section 3 then describes the corpus and the synthetic data augmentation techniques, the convolutional neural network architecture and a subsection is devoted to present the methodology. Subsequently, Section 4 points out the results and the Section 5 provides the discussion. The conclusions are presented in Section 6.

2. State of the Art

2.1. Smith–Magenis Syndrome

Smith–Magenis syndrome (SMS) is a genetically determined neurodevelopmental disorder characterized by a constellation of congenital malformations and significant behavioral and cognitive impairments [34]. It arises from a missing piece of genetic material (microdeletion) on the short arm of chromosome 17 (17p.2) in approximately 90% of cases [35,36]. In the remaining 10% [35–37], the cause stems from a mutation within a specific gene (RAI1) located in the same chromosomal region [38–40]. This gene, RAI1, is believed to play a critical role in the syndrome's development, as evidenced by its association with most of the clinical features observed in SMS patients [41,42]. The two primary diagnostic techniques employed are fluorescence in situ hybridization (FISH) and microsatellite polymerase chain reaction (PCR) within the critical SMS region [35].

Despite an estimated prevalence of 1 in 15,000–25,000 births [35,40,43], SMS remains an underdiagnosed condition. Fortunately, this delay appears to be decreasing over time. Before receiving a correct diagnosis, people with SMS often face several misdiagnoses, such as Down syndrome or autism. This long process, usually lasting years, happens because SMS is rare and not well known among physicians since it was only discovered in [44]. Delayed diagnosis of SMS has serious negative effects. The syndrome causes significant neurodevelopmental delays, leading to a severe slowdown in motor and cognitive skill development during early childhood.

Several characteristics of SMS (such as low muscle tone, delayed standing, hyporeflexia, high pain threshold, hearing impairment, intellectual disability [45,46], delayed language development [47], circadian rhythm disturbance, and behavioral disturbances [48], among others [35,45,49–52], manifest slightly in the first few months of life, making early identification a challenge [44,53]. Consequently, these early signs are often not interpreted as indicative of the underlying disorder. However, delayed diagnosis creates a substantial lag in implementing these vital clinical decisions, hindering efforts to address the marked developmental delays. This delay in diagnosis and intervention has a cascading effect, leading to a more pronounced and long-lasting impact of the syndrome's maturational impairments [30,31].

2.2. Clinical Application of Deep Learning Through the Analysis of Dysarthria

There is a compelling opportunity for research that utilizes deep learning methodologies and phonation analysis techniques in the study of neurological diseases (Parkinson's and Alzheimer's) and neurodevelopmental syndromes (SMS). A recent comprehensive review of the literature highlighted a scarcity of such research.

Voice analysis tools aid in characterizing phonation in neurological diseases like Parkinson's, where laryngeal tremors occur in 60% of cases [54]. Fluctuations in frequencies between 6 and 10 Hz signify neurodegenerative activity, along with symptoms like excessive vocal muscle tension and changes in fundamental frequency (f_0) [1,55,56]. Studies on dysarthria primarily focus on speech articulation and unintelligibility levels, utilizing acoustic indices derived from the first and second formants (F1 and F2) [57–61]. Research in this field has been extensive, with studies by Little and Chen [62,63] analyzing phonation characteristics in Parkinson's patients. Additionally, Bot and Arora [64,65], along with Iyer et al. [66], focused on voice recordings using machine learning techniques. Almeida et al. [67] found voice tasks more effective in detecting Parkinson's than speech tasks. In the realm of deep learning models, Guruler et al. [18], Ali et al. [15], and Haq et al. [17] proposed various neural-network-based approaches. Gunduz et al. [68] utilized datasets like the Parkinson's Telemonitoring Voice Data Set, while Guatelli et al. [16] are developing convolutional neural network models using mobile phone-collected voice signals. While Parkinson's remains a primary focus, there is emerging interest in Alzheimer's disease [69–72], though fewer studies have explored phonation data in this context [9,73–76].

Due to the specificity of SMS, there is a paucity of research investigating the application of deep learning techniques to vocal analysis within this population. Nevertheless, studies exploring the intersection of deep learning and speech disorders, particularly dysarthria, do not exist. Consequently, this literature review incorporates research on deep learning approaches to dysarthria, given its relevance to the broader domain of speech disorders. In [77], the authors explored a promising approach to enhance dysarthric speech recognition by augmenting the dataset with healthy speech samples. This technique could lead to improve communication tools for individuals with dysarthria. In [78], the authors discussed the challenges of recognizing disordered speech using automatic speech recognition (ASR) technologies using the UASpeech database. Despite the rapid progress in ASR for normal speech, disordered speech remains a highly challenging task. Shahamiri represents an innovative approach that combines deep learning and visual features to improve dysarthric speech recognition accuracy, benefiting patients with dysarthria [79]. The authors proposed an innovative approach that converts voice signals into visual rep-

representations called voicegrams. Joshy et al.'s study endeavored insights into using deep learning techniques and acoustic features to automatically classify dysarthria severity levels. This research work contributed to improving communication tools and therapies for individuals with dysarthria [80,81]. The former explored the automatic classification of dysarthria severity using three deep learning approaches: DNN, CNN, and LSTM with the UA-Speech and TORGO databases. The latter compared various deep learning models for classifying dysarthria severity levels using disorder-specific acoustic features using MFCCs combined with i-vectors in a DNN classifier. This study [82] focused on speech recognition for patients with hyperkinetic dysarthria disorder (RHDD) using a CTBNet model that combines convolutional neural networks (CNNs) and bidirectional gated recurrent units (Bi-GRUs) with an attention-CTC decoder to handle the temporal aspects of dysarthric speech. In [83], the authors provided a comparative analysis of pretrained deep learning models for detecting dysarthric speech using the scalogram.

Dysarthria and SMS are two distinct conditions, although both can affect communication. Dysarthria is a motor speech disorder caused by weakness, paralysis, or lack of coordination in the muscles used for speech. In contrast, SMS is a genetic disorder, which impacts physical, intellectual, and behavioral development. As mentioned before, the previously cited authors have worked on speech problems using databases of individuals with dysarthria speech, mostly in English, from multiple patients and conditions. The novelty of this work lies in the dataset involving individuals with Spanish SMS. Some of them work at the level of word or character instead of discourse, i.e., speech to text techniques. In addition, while previous research has centered on transforming audio into visual representations, i.e., voicegrams or scalograms, this research work directly processes audio signals in their raw format, placing it within the domain of sound processing. Furthermore, some studies use CNNs, a method also employed in the authors' proposal. However, in this case, the authors additionally propose the use of audio synthetic data instead of synthetic visual data, given the specific characteristics of the study population.

3. Methods and Materials

3.1. Dataset

The dataset utilized in this research comprises two distinct groups: the normative group (NG) and the SMS population (SMS). This corpus stands as a significant contribution to the field, as it contains recordings encompassing a diverse array of linguistic features. Its compilation followed a rigorous methodology for data collection. The target population for the recordings was defined as the SMS population within the Spanish Association of Smith–Magenis Syndrome (ASME), while the NG was sourced from the Public School “María Luisa Cañas” (Ciudad Real), where tutors and teachers were duly informed about the research. An explicit exclusion criterion for constituting the NG was the presence of vocal pathology, a condition verified by the school's speech therapist and confirmed by the parents. Informed consent was obtained from all tutors and parents of underage participants involved in this research. This study upholds ethical standards and does not infringe upon the rights of minors, adhering to the principles outlined in the Declaration of Helsinki by the World Medical Association in 1964 [84]. It is noteworthy to mention that SMS has a low prevalence and is often underdiagnosed. In Spain, the average age of diagnosis for this syndrome is 6.5 years [30], and ASME currently accounts for 72 diagnosed cases.

A rigorous methodology was followed for data collection. Each participant was instructed to sustain the vowel “a” for a duration of several seconds at a consistent pitch, with this protocol being iterated approximately five times per participant. Subsequently, all audio recordings underwent meticulous trimming to a standardized duration of half second, as it was determined that this temporal interval sufficed for the comprehensive extraction of all necessary features.

In summary, a total cohort of 19 participants diagnosed with SMS and 292 individuals drawn from the normative group (NG) were enrolled in the study case. Consequently,

a corpus comprising 93 audio samples from the SMS cohort and 1452 from the NG was accurately compiled.

3.2. Methods

The development of advanced techniques in image processing, collection, storage, and analysis has led to a growing adoption of these techniques for addressing class imbalance in datasets. These techniques play a crucial role in ensuring a more equitable distribution of instances across various classes. In imbalanced datasets, where one class significantly outnumbers the others in terms of the number of data points, traditional machine learning algorithms may encounter difficulty in accurately learning patterns and making predictions. This can often result in biased models that favor the majority class. The success of a neural network resides in the amount of data it is offered. The more robust the information it must analyze, the greater the specialisation and efficiency of the network are [85,86]. Exploration of oversampling and undersampling strategies aims to generate a training dataset that achieves equilibrium in class representation [87,88]. Similar to image datasets, the phenomenon of class imbalance is also prevalent within audio data samples, where distinct classes exhibit significant variations in representation [89]. Specifically, there exists a discernible majority class alongside one or more minority classes. The present study exploits aforementioned advancements to address the challenges posed by class imbalances inherent within audio datasets of this nature.

3.2.1. Oversampling

Oversampling techniques are data augmentation methods used in machine learning to address imbalanced data classification. In such scenarios, the minority class (SMS group) has significantly fewer instances than the majority class (normotypic group), leading to model bias and poor performance in identifying the minority class. The goal of oversampling is to balance the class distribution by increasing the minority class instances in the training dataset, thus improving model performance and reducing bias. This research uses several oversampling techniques: ROS, SMOTE, and variants. This manuscript considers a broad range of techniques for the synthetic generation of samples to handle highly imbalanced data, which are well described in the literature and for which software implementations are available [90,91]. Moreover, the default parameter values were used for each method. For example, the default values for the differential evolution algorithm are five neighbors, thirty clusters for cleansing, a crossover rate of 0.5 (regulates the degree of information exchange between vectors in the differential evolution process), and a similarity threshold of 0.5 (determines how similar samples must be within a cluster to consider removing redundant ones). The only modified parameter is the proportion, which is set to 0.75.

Random oversampling (ROS) focuses on the SMS class, the subset with the fewest instances, by duplicating its data points. This simple technique increases the minority class instances, potentially improving the model's ability to classify it accurately. However, replicating data points can lead to overfitting, resulting in suboptimal performance with new data as overfitting. Furthermore, the expansion of the dataset translates to an increased volume of data for the model to undergo training, consequently prolonging the training duration [92].

Synthetic minority oversampling technique (SMOTE) addresses class imbalance by generating synthetic instances for the minority class (SMS) [93]. By creating new data points similar to existing ones, SMOTE increases the minority class volume, enhancing model efficacy. However, SMOTE can introduce noise and overfitting, especially with overlapping class distributions or noisy datasets. Additional techniques may be needed to mitigate these risks. Therefore, some extra techniques should be included and they are described below.

1. Adapted density oversampling along the first principal component axis (ADOMS) addresses the challenge of class imbalance in a sophisticated manner. This method

generates synthetic examples for the minority class. Notably, these synthetic samples are strategically positioned along the first principal component axis of the local data distribution. This targeted approach ensures that the generated samples effectively augment the representation of the minority class in the training data. Consequently, ADOMS leads to improved performance by the classification model [94].

2. Assembled SMOTE distribution aims to minimize class overlap by considering the distribution of borderline minority class samples. This method creates new synthetic data points for each subgroup within the minority class. By incorporating nonborderline data points into the search for neighbors of borderline samples, especially in cases of uneven data distribution, the method ensures a pool of at least k -neighbors for comparison [95].
3. The combined cleaning and resampling (CCR) methodology adopts a holistic approach towards data preprocessing, integrating two distinct techniques: cleaning and resampling. The former facet entails the identification and subsequent elimination of noisy or irrelevant data points, with particular emphasis on those originating from the majority class. Such detailed curation aims to mitigate potential model misguidance stemming from erroneous data. The latter facilitates the oversampling of the minority class [96].
4. Cluster ensemble synthetic minority oversampling technique (CE-SMOTE) leverages cluster ensemble technology to address class imbalance within datasets. The core principle of this approach involves the introduction of a clustering consistency index. This index is employed to identify data points belonging to the minority class that reside on the periphery of clusters. Once these “cluster boundary positive samples” are identified, CE-SMOTE oversamples them to augment the original dataset and achieve a more balanced class [97].
5. Clustered synthetic minority oversampling technique (Cluster-SMOTE) combines clustering with SMOTE to improve synthetic data generation for the minority class. First, a clustering algorithm like k -means partitions the minority class data into clusters, identifying homogeneous regions. Then, SMOTE is applied within each cluster to generate synthetic data points, enhancing minority class representation. This approach creates more representative synthetic data, potentially improving model performance. However, the effectiveness of Cluster-SMOTE depends on selecting and optimizing the clustering algorithm and its parameters. Additionally, this technique involves an extra clustering step compared to standard SMOTE [98].
6. Density-based synthetic minority oversampling technique (DBSMOTE) builds on SMOTE by using density-based clustering (DBSCAN) to generate synthetic data points for the minority class. It identifies clusters of minority class data points and creates synthetic instances within these clusters, improving model performance by focusing on dense regions. However, DBSMOTE’s effectiveness depends on selecting appropriate DBSCAN parameters and may not be advantageous for uniformly distributed minority classes, where standard SMOTE might be more suitable [99].
7. Differential evolution (DE) is a prominent metaheuristic optimization algorithm that operates on a population of candidate solutions. Its ease of implementation, rapid convergence properties, and robust performance have led to its extensive adoption across various scientific and engineering disciplines. Notably, DE demonstrates particular efficacy in tackling global optimization problems within the numerical domain [100].
8. Geometric SMOTE (G-SMOTE) uses a geometric approach instead of linear interpolation. It creates a secure geometric domain (like a hyper-sphere or hyper-ellipsoid) around each minority class data point to generate synthetic data points within this domain. This method aims to produce synthetic points that are less prone to noise or outliers and more varied compared to traditional SMOTE. However, accurately defining the geometric region and its parameters adds complexity [101].

9. Gaussian SMOTE is a variant of the traditional SMOTE algorithm that uses Gaussian distribution principles to generate synthetic data points for the minority class. It extends the basic concept by incorporating Gaussian distribution attributes centered on the minority class data point, which governs the spatial displacement of the new synthetic data points. This method mitigates overfitting by introducing controlled randomness and emphasizes generating synthetic points close to the selected minority class point. However, it requires precise definition of Gaussian distribution parameters, adding complexity [102].
10. Lee's method introduces a location-aware approach for generating synthetic data with a noise-filtering mechanism. It assesses each potential synthetic data point and excludes those identified as noise, eliminating the need to predefine the number of nearest neighbors. This method also uses a rejection level metric to mitigate overfitting and noisy data generation. Synthetic instances with a rejection level of three or more are included, while those below this threshold are rejected to prevent noise [103].
11. Learning vector quantization-based synthetic minority oversampling technique (LVQ SMOTE) combines learning vector quantization (LVQ) with SMOTE to improve the generation of synthetic data points for the minority class. It involves three steps. the first one, LVQ, creates codebooks representing prototypical data points for each feature. The second one, uses these codebooks to guide the generation of synthetic data points towards regions with similar attributes to the minority class. Finally, LVQ-SMOTE then interpolates to create new synthetic data points within a "safe geometric region" around the original data point. However, defining the codebooks and their parameters adds complexity compared to standard SMOTE [104].
12. Normal distribution oversampling (NDO-Sampling) uses a random walk from existing minority class data points to generate synthetic samples. This method ensures that the synthetic samples closely match the original minority class data's expected value and squared deviation. It mitigates intraclass similarity and addresses data imbalance efficiently without needing to find the k-nearest neighbors. This approach helps avoid overfitting and generates new data points that resemble the original minority class data without complex neighbor searching algorithms [105].
13. Noncooperative game theory for filtering oversampled data (NEATER), introduced in [106], exploits game theory principles for data refinement. This method discards data points within the oversampled set that exhibit a heightened likelihood of belonging to the opposing class. The likelihood of class membership is determined through an analysis of each instance's neighborhood within the data. However, this process can introduce noise or redundant data. NEATER aims to filter out these unwanted elements [106].
14. Proximity weighted synthetic oversampling technique (ProWSyn) enhances SMOTE by using proximity weighting, assigning higher weights to minority class data points near the decision boundary. This prioritizes generating synthetic data in critical regions, improving the model's ability to recognize the minority class. By focusing on the decision boundary, ProWSyn can better capture the minority class's intricacies and improve classification accuracy. Additionally, it helps avoid generating synthetic points in noisy regions, unlike traditional SMOTE. The effectiveness of ProWSyn depends on the method used to calculate and apply proximity weights, requiring careful calibration for optimal performance [107].
15. Synthetic minority oversampling based on data density (SMOBD) employs a range of techniques. It includes traditional (i) support vector machines (SVMs) where SMOBD leverages this information to generate new data points that line the decision boundary, focusing on areas where the minority class is scarce; (ii) SMOTE combined with edited nearest neighbors (SMOTE-ENN), which helps to remove these noisy data points before applying SMOTE; and (iii) SMOTE with varying costs for SVM (SDC) that incorporates class imbalance information by assigning different costs to misclassifying data points during the SVM training process [108].

16. Supervised SMOTE. The primary objective of supervised oversampling (supervised SMOTE) lies in achieving a more balanced distribution within datasets. This objective is accomplished by synthesizing additional data samples that belong to the minority class. The process of generating these synthetic samples is guided by the existing data points, hence the term “supervised” [109].

3.2.2. Undersampling

Undersampling is a method in machine learning used to address imbalanced data classification. It reduces the size of the majority class to balance the class distribution, allowing the model to focus on the minority class. In this research, the minority class is the SMS individual group, and the majority class is the normotypic class group [110]. Two strategies are proposed: edited nearest neighbors (NearMiss), which removes majority class samples similar to minority class examples, and random undersampling (RUS), which selectively removes instances from the majority class. The former approach removes samples of the normotypic class that are very similar (nearest neighbors) to minority class examples. This aims to prevent removing informative majority class examples that lie close to the decision boundary [111]. The latter, in the domain of imbalanced learning, presents a technique for addressing class imbalance within datasets, particularly those employed for classification tasks. This scenario arises when one class significantly outnumbers the others. RUS achieves a more balanced class distribution by selectively removing or sub-sampling instances from the majority class. This approach mitigates the dominance of the majority class, thereby enhancing the classifier’s ability to learn effectively from the minority class data.

3.3. Neuronal Network (CNN)

The neural network architecture depicted in Figure 1 showcases its diverse layers, shapes, and hyperparameters.

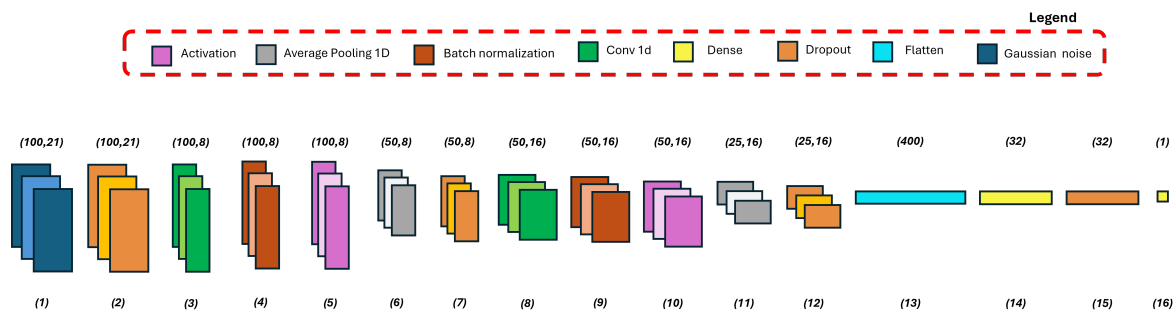


Figure 1. Architecture of the convolutional neural network (CNN) proposed for the current research work. (1) Gaussian noise (stddev = 0.01); (2) dropout (0.1); (3) Conv1d (filters = 8, kernel_size = 3, padding = ‘same’); (4) batch normalization; (5) activation (Relu); (6) average pooling 1D (pool_size = 4, strides = 2, padding = ‘same’); (7) dropout (0.1); (8) Conv1d (filters = 16, padding = ‘same’); (9) batch normalization; (10) activation (Relu); (11) average pooling 1D (pool_size = 4, strides = 2, padding = ‘same’); (12) dropout (0.1); (13) flatten; (14) dense (activation=‘relu’); (15) dropout (0.1); (16) dense (activation = ‘relu’).

The architectural journey commences with the input layer. The Gaussian noise layer introduces random Gaussian noise to the input data with a standard deviation of 0.01. This stochastic process adds variability to the input, serving to prevent overfitting during training. The dropout layer is a regularization technique commonly applied in neural networks. During training, a fraction of randomly selected neurons within the network are temporarily “dropped out” or ignored. This dropout process helps prevent the network from relying too heavily on specific neurons or features, thus reducing the risk of overfitting. A dropout rate of 10% is applied during each of the four steps where this model is utilized.

Subsequently, the network progresses to the convolutional layers, the fundamental building blocks of a CNN. These layers employ a multitude of filters, mathematically termed kernels that traverse the audio in a systematic fashion. The 1D convolutional layer performs convolutions on one-dimensional input data. Utilizing learnable filters or kernels, this layer extracts relevant features from the input signal, capturing dependencies and patterns. By sliding the kernel across the input sequence and computing dot products, the layer generates feature maps that represent abstracted representations of the input data. This process enables the network to effectively learn hierarchical features essential for various tasks, such as sequence modeling and time series analysis. The convolutional neural network commences operation with an initial frame count of one hundred. Each frame encapsulates a feature vector comprised of twenty MFCCs alongside a CPP feature. The first iteration leverages a convolutional layer configured with eight filters. These filters are applied by a set of three kernels. The application of these filters adheres to a padding convention, ensuring consistency in the resulting feature map dimensions. The subsequent iteration processes the output from the prior phases. This iteration employs a convolutional layer with sixteen filters. The kernel quantity and padding convention remain identical to those utilized in the first iteration.

The batch normalization layer is a normalization technique applied to the activations of a neural network's hidden layers. By normalizing the activations within each mini-batch during training, batch normalization helps mitigate internal covariate shift and accelerates the training process. It is applied twice in the CNN sequence.

The activation layer plays a crucial role. It utilizes a nonlinear function to introduce complexity, enabling the network to learn nonlinear relationships within the data. Among these nonlinear functions, the rectified linear unit (ReLU) introduces nonlinearity to the network by thresholding the input at zero, effectively replacing negative values with zeros while leaving positive values unchanged. Like above, this process is repeated twice. Following each activation function, an average pooling layer is applied. This operation reduces the spatial dimensionality of the feature maps. The input size to the pooling layer is halved compared to the preceding layer's output. The number of filters used in the convolutional layers, as previously established, remains at 8 and 16 for the first and second iterations, respectively.

The flattening layer is executed after the third dropout step. It takes the multidimensional output from the convolutional and pooling layers and converts it into a single, one-dimensional vector. In this particular case, the aforementioned layer produces a feature vector whose dimensionality is 400. This dimensionality arises from the concatenation of two sources: the initial set of twenty MFCCs and the output from the preceding convolutional layer, which employs sixteen filters.

Finally, the first dense layer employs an ReLU activation function. The final dense layer, responsible for generating the network's output, utilizes a sigmoid activation function. This function maps the output values between normotypic and SMS individual, making it suitable for tasks like binary classification, where the network needs to predict the probability of an instance belonging to a particular class.

3.4. Metrics

In this context of SMS classification, a true positive (TP) refers to an SMS individual that is correctly identified as an SMS. Conversely, a true negative (TN) signifies a non-SMS participant that is accurately classified as such. However, classification errors can occur. A false positive (FP) denotes a non-SMS individual mistakenly categorized as an SMS, and a false negative (FN) represents an SMS individual incorrectly classified as a non-SMS. Understanding these terms is crucial for evaluating the performance of SMS classification algorithms.

The performance of a classifier in binary classification can be evaluated using three key metrics: sensitivity, specificity, and accuracy.

Sensitivity, known as the true positive rate, reflects the proportion of positive cases that the model correctly identifies. It essentially calculates the ratio between true positives (correctly classified positive examples) and all the samples that were originally labeled as positive.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

Specificity, however, focuses on the model's ability to correctly classify negative cases. It is calculated as the ratio between true negatives (correctly classified negative examples) and all the samples initially labeled as negative.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

Balance accuracy, while seemingly straightforward, can be misleading in imbalanced datasets. It is simply the sum of true positives and true negatives divided by 2.

$$\text{Balance Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (3)$$

3.5. Methodology

Figure 2 depicts the application's comprehensive workflow, which can be broken down into seven distinct phases. This research work utilizes three study cases. Study cases 1 and 2 delve into specific aspects of a seven-stage process, focusing on all phases except phases 4 and 5, respectively. In contrast, study case 3 offers a comprehensive examination by encompassing all seven stages. In this final study case, the computation time, which was the longest, was approximately 9 h and was executed on an instance of a supercluster with the following features: 32 VCPU, 60.5 GB of RAM, and 16 GB of virtual hard disk.

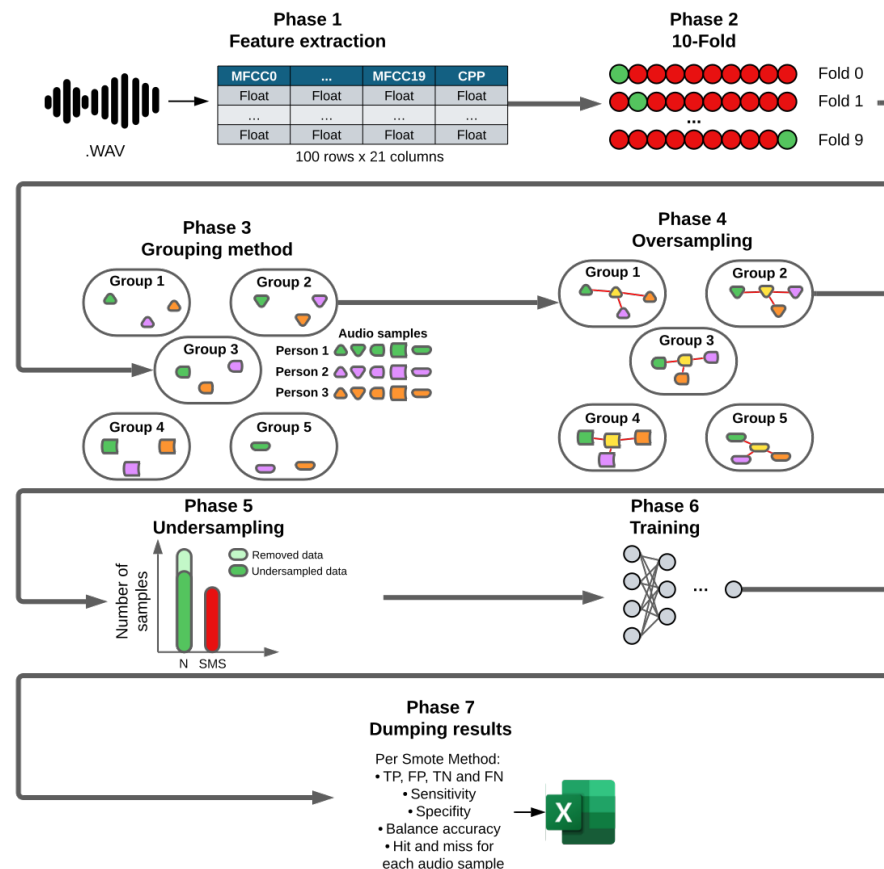


Figure 2. Main phases of the application.

- Phase 1: Feature extraction involves the automatic extraction of mel-frequency cepstral coefficients (MFCCs) [112] and cepstral peak prominence (CPP) features from audio files in WAV format. MFCCs are effective for tasks like speaker and speech recognition, while CPPs represent the most prominent spectral peaks. The extracted features are transformed into a 100×21 matrix, with the first 20 columns holding MFCC coefficients and the 21st column storing the CPP value. This data structure provides valuable insights into the audio signal's characteristics. For a visual representation of this data structure, please refer to Figure 2 phase one.
- Phase 2: Stratified K-fold involves separating data into distinct folds before adding synthetic data to prevent data leakage, which can compromise model evaluation. Data leakage occurs when the model accesses test set information during training, leading to inflated performance. To avoid this, a 10-fold cross-validation approach is used, partitioning data into ten equally sized folds at the individual level. This ensures that all audio samples from a single participant are in the same fold, preventing the model from encountering information about a participant during training that it might see during validation. This approach provides a more reliable evaluation of the model's ability to generalize to unseen data.
- Phase 3: Grouping method expands and diversifies datasets in machine learning but can introduce redundant information when synthetic samples closely resemble existing data. To address this, a stratified oversampling technique focusing on redundancy reduction is proposed. This process is applied to both the SMS and control groups, ensuring dataset enrichment while minimizing redundancy. If an individual has more than five original samples, the extras are excluded from oversampling but are added to the training set later. This ensures balanced representation and reduces redundancy. A nonstratified approach, like SMOTE, could introduce bias by selecting nearest neighbors from the same individual, leading to redundant information. Stratified oversampling avoids this by ensuring that each fold during cross-validation contains no more than one original sample per individual, promoting a diverse pool of nearest neighbors and enriching the dataset while mitigating redundancy.
- Phase 4: Oversampling. During this stage, the focus shifts to augmenting synthetic data within each group established earlier. The goal is to create a comprehensive and diverse final training dataset, enhancing the robustness of analyses and model performance. Tailored data augmentation methods are applied to each group. Maintaining a balanced sample distribution is crucial, with both original and augmented samples consistently at 75%. This balance prevents noise or redundancy, preserving the integrity of analyses. Initially, the normotypic group had 1302 samples, and the SMS group had 83. After augmentation, the dataset totaled 996 samples (in study case 1 and study case 3), highlighting the process's effectiveness in increasing volume and diversity. All groups are then combined into a unified dataset. Subsequent undersampling is conducted globally, reducing the risk of discarding relevant samples.
- Phase 5: Undersampling. This study used the second variant of the NearMiss algorithm to address class imbalance by removing majority class (normotypic) samples closest in mean distance to the farthest minority class samples. This approach ensures a balanced representation of both classes, crucial for effective model training. The NearMiss algorithm balances reducing class imbalance with minimizing the loss of informative data. In this study, 90% of the majority class samples were removed, retaining 10% closest to the farthest minority class samples. The final training set had 1106 normotypic and 996 SMS samples in study case 3.
- Phase 6: Model training. To ensure robust performance evaluation, a 10-fold cross-validation approach is used, partitioning data into ten equal folds. The CNN trains for twenty epochs per fold, iteratively refining its parameters. This process is repeated for all ten folds, ensuring diverse data evaluation and mitigating biases. To address imbalanced datasets, sixteen SMOTE variations generate synthetic minority class data, balancing the class distribution. The training protocol is executed for each SMOTE

variation, resulting in 160 training iterations (10 folds $\times$ 16 SMOTE variations). This comprehensive approach ensures a statistically reliable evaluation of the methodologies.

- Phase 7: Dumping results. Due to the numerous methods used, outcomes are automatically dumped into a spreadsheet. Each SMOTE method has its own Microsoft Excel spreadsheet with two tables: one for audio data (including speaker and audio identifiers, hits, and misses) and another for the confusion matrix and performance metrics (sensitivity, specificity, balanced accuracy). This approach allows comprehensive analysis and comparison across iterations, aiding decision making on SMOTE methods. A structured results repository is created for easy comparison of methodologies. It includes a second table identifying the most and least successful audio samples, offering insights into factors affecting model performance.

4. Results

Three distinct study cases were conducted to comprehensively assess the impact of both undersampling and oversampling methodologies on the performance of the model. Through research work and analysis, these studies aimed to elucidate the nuanced effects and efficacy of various sampling techniques in enhancing model performance across diverse datasets and scenarios.

The first study case (SC1) employed two well-known oversampling methods: traditional SMOTE and random oversampling (ROS). It is important to highlight that the preprocessing and training procedures remained consistent with those depicted in Figure 2 excluding the undersampling phase. In the second study case (SC2), two renowned undersampling methods, NearMiss and random undersampling (RUS), were employed. Similar to the previous case, the methodology adhered to the model outlined in Figure 2, with the exclusion of the oversampling phase, focusing solely on the utilization of undersampling techniques. Finally, in the third study case (SC3), it is proposed to integrate oversampling and undersampling techniques in accordance with the framework delineated in Figure 2. The oversampling and undersampling methodologies employed correspond to those explained in Sections 3.2.1 and 3.2.2. Among the techniques detailed in Section 3.2.2, the NearMiss method was selected for implementation. This choice was motivated by the study's purpose to investigate alternatives to conventional SMOTE and to evaluate the advantages of their fusion with an undersampling approach, particularly in scenarios characterized by a pronounced imbalance rate.

4.1. Study Case 1—SMOTE and ROS

The oversampling technique specifically targets the minority class, which possesses a lower number of instances. Both methods, SMOTE and ROS, aimed to address class imbalance by increasing the number of data points in the SMS class, the minority one. This helps to create a more balanced dataset where the minority class has a greater influence on the model's learning process.

Random oversampling operates by duplicating or synthesizing minority class examples, effectively increasing their representation within the dataset and aiming to achieve a more balanced class distribution. On the contrary, SMOTE creates synthetic data points in the feature space between existing minority class data points. Consequently, there was a deliberate effort to evaluate the performance and advancements of both methodologies using the dataset outlined earlier. To accomplish this objective, each methodology was executed iteratively ten times, and, subsequently, the mean outcomes along with their corresponding standard deviations were meticulously documented, as illustrated in Table 1. The results obtained for the balance accuracy were 88.5% and 88% for SMOTE and ROS, respectively. An analysis of the data revealed a close correspondence between specificity and sensitivity values. Both methods achieved an accuracy of 80% in correctly classifying SMS samples. However, SMOTE exhibited a slight advantage in correctly identifying normotypic samples, achieving a value of 96% compared to ROS's 95%. Finally, it is noteworthy that both methods correctly classified 75 SMS samples. However, a distinction emerged in their

performance with normotypic samples. SMOTE achieved a higher number of correctly classified normotypic samples (1396) compared to ROS (1382).

Table 1. Representation of all averages and standard deviations obtained after the 10 iterations of each of the oversampling methods, SMOTE and ROS, in the SC1.

Metric Method	SMOTE	ROS
TP	75.1 ± 3.107	75.2 ± 4.442
FN	17.9 ± 3.107	17.8 ± 4.442
TN	1396.9 ± 14.571	1382.1 ± 30.468
FP	55.1 ± 14.571	69.9 ± 30.468
Sensitivity	0.808 ± 0.033	0.809 ± 0.048
Specificity	0.962 ± 0.010	0.952 ± 0.021
Bal. Accuracy	0.885 ± 0.017	0.880 ± 0.027

4.2. Study Case 2—NearMiss and RUS

This study case focused on NearMiss undersampling, variant 2, and random undersampling (RUS) methods, techniques used to address class imbalance in classification tasks. Both techniques belonged to the category of undersampling techniques. They achieve class balancing by removing data points from the majority class. Class imbalance, the normative group, had significantly more data points compared to the minority class, SMS group. This imbalance could lead to models that are biased towards the majority class and perform poorly on the minority class.

Although the NearMiss method increased complexity, the results obtained were slightly better than RUS. NearMiss requires calculating distances between data points, making it computationally more expensive than random undersampling. Furthermore, by considering data point similarity, NearMiss could potentially retain more informative data from the majority class compared to random undersampling, as it is shown in Table 2. While both RUS and NearMiss undersampling techniques achieved high specificity, with RUS reaching 90% and NearMiss reaching 89%, a closer examination revealed a key advantage for NearMiss. This advantage lies in its superior sensitivity, which stood at approximately 4% higher compared to RUS. This difference in sensitivity translates to a slightly better balance accuracy ratio for NearMiss. The outcomes related to the balance accuracy were 78.4% and 76.8%, respectively.

Table 2. Representation of all averages and standard deviations obtained after the 10 iterations of each of the undersampling methods, NearMiss and RUS, in the SC2.

Metric Method	NearMiss	RUS
TP	62.7 ± 3.466	58.6 ± 4.904
FN	30.3 ± 3.466	34.4 ± 4.904
TN	1297.8 ± 74.207	1314.4 ± 91.951
FP	154.2 ± 74.207	137.6 ± 91.951
Sensitivity	0.674 ± 0.037	0.630 ± 0.053
Specificity	0.894 ± 0.051	0.905 ± 0.063
Bal. Accuracy	0.784 ± 0.024	0.768 ± 0.033

4.3. Study Case 3—Oversampling and Undersampling

In the final scenario proposed, a novel approach was adopted, combining oversampling with undersampling techniques. This study case thoroughly examined various SMOTE methods outlined in the methods section, each integrated with the NearMiss undersampling technique. By analyzing these combined approaches, this study aimed to provide a comprehensive understanding of their effectiveness in mitigating class imbalance. Each oversampling technique associated with the aforementioned NearMiss method was executed ten times. Following each iteration, the above performance metrics were

calculated. Finally, the average and standard deviations are presented for each combination. As illustrated in Table 3, the methods exhibiting the most favorable outcomes are consistently the ones with the highest performance metrics (colored in green): DB-SMOTE, CCR, and NDO_sampling, which obtained the best outcome balance accuracy of 91.2%. Conversely, at the opposite end of the spectrum, the methods yielding the lowest results consistently manifested shortcomings in performance (colored in red): SMODB and DE_oversampling with 69.3% and 70.8% balance accuracy, respectively.

The results corresponding to the methods with the best performance present sensitivity and specificity values greater than 0.80 and 0.95, respectively. The best sensitivity score corresponds to NDO_sampling with 0.870 ± 0.040 . For its part, the best specificity is provided by the G-SMOTE method with 0.976 ± 0.007 . Although none of the analyzed techniques offers a simultaneous improvement in both measures, the sensitivity decrease experienced by the last one (0.776 ± 0.042) is noticeably greater than can be seen for the specificity in NDO_sampling (0.954 ± 0.013). Finally, as can be seen, the worse performing algorithms (marked in red in Table 3) present a noticeable reduction for both scores at the same time.

Table 3. Representation of all averages and standard deviations obtained after the 10 runs of each oversampling method together with the NearMiss method in the SC3. The total number of oversampling methods is 16, as shown in the first column.

Methods Metrics	TP	FN	TN	FP	Sensitivity	Specificity	Bal. Accuracy
ADOMS	72.4 ± 4.195	20.6 ± 4.195	1397.3 ± 17.366	54.7 ± 17.366	0.778 ± 0.045	0.962 ± 0.012	0.870 ± 0.022
Assembled SMOTE	71.6 ± 3.950	21.4 ± 3.950	1414.7 ± 11.833	37.3 ± 11.833	0.770 ± 0.042	0.974 ± 0.008	0.872 ± 0.021
CCR	76.8 ± 3.084	16.2 ± 3.084	1387.7 ± 33.012	64.3 ± 33.012	0.826 ± 0.033	0.956 ± 0.023	0.891 ± 0.015
CE SMOTE	72.6 ± 4.274	20.4 ± 4.274	1407.0 ± 8.083	45.0 ± 8.083	0.781 ± 0.046	0.969 ± 0.006	0.875 ± 0.023
Cluster SMOTE	74.0 ± 4.595	19.0 ± 4.595	1396.5 ± 15.580	55.5 ± 15.580	0.796 ± 0.049	0.962 ± 0.011	0.879 ± 0.025
DBSMOTE	75.0 ± 3.712	18.0 ± 3.712	1395.2 ± 16.665	56.8 ± 16.665	0.806 ± 0.040	0.961 ± 0.011	0.884 ± 0.018
DE oversampling	61.4 ± 3.893	31.6 ± 3.893	1096.1 ± 71.286	355.9 ± 71.286	0.660 ± 0.042	0.755 ± 0.049	0.708 ± 0.023
G-SMOTE	72.2 ± 3.938	20.8 ± 3.938	1416.6 ± 10.298	35.4 ± 10.298	0.776 ± 0.042	0.976 ± 0.007	0.876 ± 0.020
Gaussian SMOTE	61.3 ± 3.592	31.7 ± 3.592	1376.6 ± 38.549	75.4 ± 38.549	0.659 ± 0.039	0.948 ± 0.027	0.804 ± 0.019
Lee	70.0 ± 2.055	23.0 ± 2.055	1408.4 ± 21.593	43.6 ± 21.593	0.753 ± 0.022	0.970 ± 0.015	0.861 ± 0.013
LVQ_SMOTE	77.3 ± 4.084	15.7 ± 4.084	1312.1 ± 53.272	139.9 ± 53.272	0.831 ± 0.044	0.904 ± 0.037	0.867 ± 0.024
NDO Sampling	80.9 ± 3.725	12.1 ± 3.725	1384.8 ± 18.220	67.2 ± 18.220	0.870 ± 0.040	0.954 ± 0.013	0.912 ± 0.019
NEATER	65.0 ± 4.447	28.0 ± 4.447	1395.6 ± 13.737	56.4 ± 13.737	0.699 ± 0.048	0.961 ± 0.009	0.830 ± 0.022
ProWSyn	70.9 ± 3.178	22.1 ± 3.178	1400.7 ± 15.924	51.3 ± 15.924	0.762 ± 0.034	0.965 ± 0.011	0.864 ± 0.020
SMODB	67.1 ± 6.208	25.9 ± 6.208	963.8 ± 74.754	488.2 ± 74.754	0.722 ± 0.067	0.664 ± 0.051	0.693 ± 0.030
Supervised SMOTE	73.3 ± 3.498	19.7 ± 3.498	1408.6 ± 27.965	43.4 ± 27.965	0.788 ± 0.038	0.970 ± 0.019	0.879 ± 0.012

5. Discussion

Table 4 shows the research works reviewed addressing different aspects of improving automatic speech recognition (ASR) and dysarthria severity classification through deep learning techniques. In [82], the model achieved a character error rate (CER) of 15% and a word error rate (WER) of 59% on a dataset of 2797 recordings, surpassing other models like 4xLSTM and 3xCNN-3xLSTM. Similarly, ref. [81] explored severity classification using MFCC combined with i-vectors in a DNN, achieving 93.97% accuracy in speaker-dependent scenarios. However, the performance dropped to 49.22% in speaker-independent cases, highlighting the challenge of variability in dysarthric speech. These studies emphasize the potential of DL in assistive technologies, targeting either improved speech recognition or severity classification for dysarthric patients.

Table 4. A comparative analysis of prior studies on dysarthria, detailing their datasets, features, and employed deep learning models.

Authors	Datasets	Features	Deep Learning Models	Accuracy Rate
Hashan et al. [82]	2797 Russian recordings of RHDD speech	MFCC	CTBNet, 4xLSTM, 3xCNN + 3xLSTM	CER: 15%, WER: 59%, CER: 36%, WER: 66%, CER: 37%, WER: 63%
Joshy and Rajan [81]	UA-Speech y TORGO	MFCC, CQCC, prosodic, glottal, and articulatory features, i-vectors	DNN, CNN, GRU, LSTM, i-vectors	Speaker-Dependent: 93.97% UA-Speech/ Speaker-Independent: 49.22% UA-Speech
Liu et al. [78]	UASpeech, CUDYS	MFCC, Mel-scale filter banks, delta features, pitch	DNN, TDNN, BLSTM, CTC, LAS, Pychain TDNN	25.21% WER
Shahamiri [79]	UA-Speech	Voicegrams, synthetic dysarthric speech	2D CNN	64.71% WRA
Joshy and Rajan [80]	UA-Speech, TORGO	MFCC y derivates (39 characteristics)	DNN, CNN, LSTM	TORGO: 96.18%, UA-Speech: 93.96.94 (AlexNet), 97.54 (GoogLeNet), 98.21 (ResNet 50), 82.06 (VGGish), 97.88 (YAMNet)
Shanmugapriya and Mohan [83]	Custom dataset of dysarthric speech	Scalograms of dysarthric speech signals	Pretrained CNNs (AlexNet, GoogLeNet, ResNet 50, VGGish, YAMNet)	98.21 (ResNet 50), 82.06 (VGGish), 97.88 (YAMNet)
This work	Custom dataset of Spanish SMS and normatives	MFCC, CPP	CNN, Oversampling and undersampling techniques	91.2% (NDO Sampling), 89.1% (CCR)

Other studies, like [78], focus on innovative ASR techniques, with the former utilizing automated neural architecture search and visual recognition to achieve a WER of 25.21%, and the latter introducing voicegrams to achieve 64.71% word recognition accuracy. In [80], the authors found that CNN and DNN models significantly outperformed traditional approaches, reaching 96.18% accuracy on the TORGO database and 93.24% on UA-Speech. Lastly, an evaluation of pretrained CNN models, with ResNet 50 achieving the highest accuracy of 98.21%, was conducted in [83]. Collectively, these works highlight the growing effectiveness of DL models for dysarthric speech recognition and severity assessment, driving progress in assistive technologies for individuals with speech disabilities.

This study utilized an original dataset comprising voice recordings from Spanish-speaking patients diagnosed with SMS, collected from associations and schools. Due to the scarcity of Spanish-speaking SMS patient samples, various oversampling and undersampling techniques were implemented to balance the datasets. By training a CNN and extracting acoustic features such as MFCC and CPP, a highly competitive classification rate of nearly 92% was achieved. These results demonstrate the potential of deep learning techniques for detecting SMS through voice analysis, even in scenarios with exiguous data. A discussion of the three case studies conducted is presented below.

In the SC1, SMOTE generates synthetic data points within the feature space by interpolating between existing minority class data points. Furthermore, SMOTE endeavors to grasp the fundamental data distribution and potentially introduce fresh insights to the minority class. Conversely, ROS duplicates existing minority class data points with replacement, effectively augmenting the representation of the minority class within the dataset. Additionally, it merely duplicates existing data points, which could result in overfitting and heightened variability. The aforementioned distinctions between these techniques can manifest in subtle variations within the results, as evidenced by Table 1. However, when applied to the current dataset, these variations are not statistically significant.

In the SC2, NearMiss and RUS diverge in their approaches to selecting majority class instances for retention or removal when addressing class imbalance. NearMiss prioritizes retaining instances closest to the decision boundary, aiming to preserve informative examples that may influence classification. Conversely, RUS employs a random selection process,

simply removing majority class instances until a balanced class distribution is achieved. NearMiss prioritizes retaining normative class instances closest in distance (similarity) to SMS class instances. This rationale aims to preserve informative examples residing near the decision boundary, as they hold a higher risk of misclassification. Nevertheless, RUS employs a random selection process, which removes instances from the normative class until a predetermined level of balance between classes is achieved. This method does not take into account the proximity of instances to the decision boundary.

An interesting observation emerges when comparing the performance of the NearMiss and RUS methods on SMS classification. The NearMiss method demonstrates a clear strength in identifying actual SMS individuals. It successfully classifies 62 out of 93 SMS samples ($\approx 67\%$), exceeding the 58 correctly identified by the RUS method. However, this advantage comes with a trade-off. The NearMiss method exhibits a slight weakness in correctly classifying normal, non-SMS samples. It identifies 1297 normotypic samples as such, while the RUS method achieves a slightly higher accuracy in this category, correctly classifying 1314 normotypic samples. This comparison highlights the importance of considering both aspects of classification performance—the ability to identify true SMS samples and to avoid misclassifying normal samples—when evaluating these methods.

The optimal choice between NearMiss and RUS hinges on the specific characteristics of the dataset. If preserving information within the majority class is crucial, as is the case in this dataset, NearMiss may be preferable due to its targeted selection. As demonstrated in Table 2, the aforementioned distinctions between the NearMiss method and RUS translate to marginally superior performance in terms of balance accuracy for the NearMiss method. Consequently, NearMiss was chosen as the preferred technique for application within study case 3.

Lastly, the findings from study case 3, shown in Table 3, can be categorized into three groups. Firstly, the least effective methods include SMODB and DE_oversampling, with accuracy values of 69.3% and 70.8%, respectively. These findings underscore the importance of methodological selection and optimization in achieving desired outcomes, emphasizing the need for further exploration and refinement in algorithmic approaches. Secondly, methods falling within a moderate accuracy range of 80.4% to 87.9% form the second group. Within this range ($\approx 7.5\%$), six methods achieve an accuracy of 87%, three methods attain 86%, one method reaches 83%, and another scores 80.4%. The best methods in this group are Supervised_SMOTE and Cluster_SMOTE. Lastly, the final group comprises methods that have demonstrated the most favorable outcomes. Specifically, the DBSMOTE, CCR, and NDO_Sampling methods achieved accuracy scores of 88.4%, 89.1%, and 91.2%, respectively. These methods not only showcase superior balance accuracy but also demonstrate robustness and efficacy in addressing the given problem.

One group of SMOTE variants exhibits high specificity but lower sensitivity. In other words, these methods achieve highly accurate classification in normative cases, at the expense of performance in identifying SMS cases. This group includes G_SMOTE, Assembled_SMOTE, Supervised_SMOTE, and Lee, all of which demonstrate specificity values exceeding 97%. Conversely, another group of methods achieves superior sensitivity, meaning they more effectively identify SMS cases of the minority class, but this success comes at the cost of correctly classifying majority class cases. LVQ_SMOTE and DBSMOTE are prime examples within this group, with sensitivity values of 83% and 80%, respectively. Figure 3 illustrates the interplay among sensitivity, specificity, and the equilibrium of balance accuracy, wherein sensitivity denotes the proportion of properly classified SMS cases, and specificity represents the accurate classification of normotypic individuals. The graphical depiction is organized clockwise, aligning with descending order based on the balance accuracy rates. Notably, the NDO_sampling method attains the highest value, while the SMODB method exhibits the lowest value in this continuum.

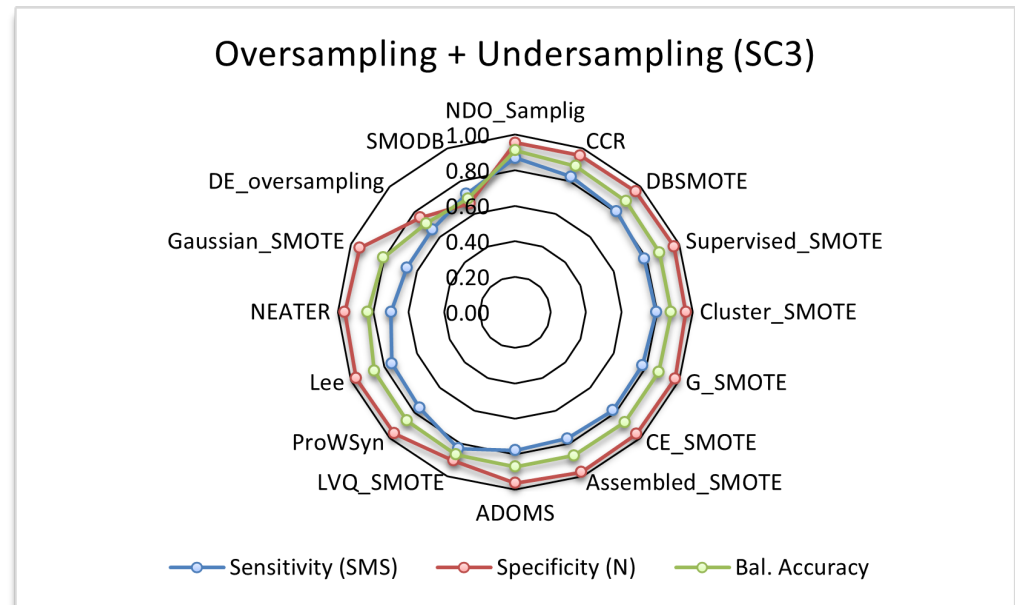


Figure 3. SC3: Outcomes of the sensitivity, specificity, and balance accuracy of the 16 SMOTE variants run in conjunction with its NearMiss.

NDO_Sampling exhibits exceptional specificity, consistently exceeding 93% across the ten iterations. The highest recorded specificity value reached 97%. These data translate to an average specificity of approximately 95%, indicating a low rate of false negatives. In simpler terms, the model demonstrates a high accuracy rate in excluding individuals who do not have the syndrome. It is noteworthy that the sensitivity values are comparatively lower than the specificity values. The average sensitivity across iterations stands at 87%, with the highest achieving 94% and the lowest at 81%. Consequently, the model occasionally fails to detect true positives with the same level of accuracy as it identifies true negatives, as indicated by its specificity. In essence, while the test demonstrates proficiency in identifying individuals who truly have the disease, its sensitivity falls short of its specificity. The CCR method demonstrates a performance profile similar to its predecessor. Remarkably, specificity values consistently exceed sensitivity values throughout the range of measurements. Specificity values range from 98% to 91%, while sensitivity values range from 87% to 72%. This pattern suggests a stronger proficiency in identifying true negatives compared to true positives across various evaluation metrics.

An examination of the data from SC1 and the results obtained in SC3 reveals minimal discrepancies between the best-performing methods. The most successful instance of SC1, utilizing SMOTE, achieved an accuracy of 88.5%, while the best case of SC3, employing NDO_Sampling, yielded an accuracy of 91.2%. This translates to a difference of less than 3%. Furthermore, the performance of the SMOTE method (88.5%) surpasses the average accuracy (84.5%) achieved by the combined set of SMOTE variants and the NearMiss undersampling method. It is also comparable to the performance of most individual SMOTE variants (86.0%). An evaluation of the sixteen variants explored within the scope of this investigation yielded only three that surpassed the performance achieved by the SMOTE technique employed in SC1. This finding necessitates a thorough examination of the cost–benefit analysis associated with these variants. It is crucial to determine whether the potential benefits garnered from these more intricate approaches outweigh the additional computational resources, methodological complexity, and time investment required for their implementation, compared to the well-established and computationally efficient SMOTE technique.

The current focus on this research work prioritizes the robustness and resilience of employed methodologies and metrics. This emphasis ensures the reliability of these methods even in nonideal circumstances, potentially at the expense of achieving the absolute highest

performance metrics. While sacrificing peak performance, true positive rate, may seem counterintuitive, it fosters a more sustainable and trustworthy approach. This trade-off between peak performance and robustness is a critical consideration when selecting and evaluating oversampling and undersampling methodologies. Henceforth, a custom metric is employed, which prioritizes the classification accuracy of the target class (SMS) over the high classification rates achieved by the neural network.

Despite the promising results, this study has some limitations. The relatively small sample size and its nature may have influenced the outcomes. Additionally, while the synthetic data model used was suitable for analyzing this data, it could be improved to capture all possible relationships between the studied variables. Although the number of speakers with SMS is small, they represent a very high percentage of all subjects diagnosed with this disease in Spain at the time of the recordings. In fact, this number constitutes approximately 20% of the total number of people with SMS in the Spanish Smith–Magenis Syndrome Association (ASME).

This article explores how the generation of synthetic samples contributes to improving the neural network training process and its influence on the subsequent testing phase, where the risk of overfitting is extremely relevant due to the notable scarcity of data for one of the classes (subjects with SMS). While the conclusions of this work, regarding which synthetic data generation methods present the best performance, may vary when applied to other similar problems, they constitute a good starting point for the study of other population groups.

6. Conclusions

This groundbreaking research utilizes a neural network model, analyzing phonation data to identify a rare genetic syndrome. Furthermore, to address the challenge of imbalanced datasets, the research work incorporated sixteen distinct variations of the SMOTE. Each SMOTE variation was employed to generate synthetic data points for the minority class, effectively balancing the class distribution within the dataset. The study achieved high accuracy in detecting the syndrome (SMS) by employing synthetically generated data (through oversampling and undersampling) alongside a CNN. This manuscript highlights the potential of voice analysis, using only two types of acoustic features: MFCC and CPP, and deep learning techniques as a noninvasive and valuable tool. This finding holds promise for improving the identification of this complex syndrome and potentially other rare diseases.

This research work presents three study cases. The first study case investigates the use of two standard oversampling techniques (SMOTE and ROS). Both techniques achieved similar results, with a classification rate of 88%. The second study case focuses on the application of two undersampling techniques (NearMiss and RUS). This approach resulted in a significant decline in performance compared to the previous study case, with NearMiss achieving only a 78% classification rate. Finally, a comprehensive study case was conducted, employing 16 oversampling techniques and one undersampling technique. This study case yielded the highest accuracy rate, with NDO_sampling and NearMiss achieving approximately 92% accuracy. It should be highlighted that the overarching objective of this study centers on determining the impact of oversampling and undersampling techniques on the accuracy rate within imbalanced datasets.

This elaborate work not only presents a significant advancement in the classification of patients with SMS through the analysis of voice samples but also paves the way for exploring similar methodologies in the classification and study of other rare genetic syndromes. Conditions such as Williams syndrome, Prader–Willi syndrome, Fragile X syndrome, and Angelman syndrome stand as examples of rare genetic disorders that could potentially benefit from similar voice-based approaches. By extending the scope of research to encompass these conditions, the broader research community can envisage new avenues for early detection, classification, and understanding of these complex disorders. Furthermore, this work can be extended by exploring hybrid models combining voice and image

analysis. The authors aim to determine if adding visual information can enhance current models, especially for tasks requiring deeper contextual understanding.

To sum up, this research work demonstrates the feasibility of using CNNs in conjunction with data augmentation techniques to develop an accurate classification system for SMS. The obtained results contribute to the field of computer-aided diagnosis and open new avenues for the development of voice-based screening tools. While the results are promising, further studies with larger and more diverse datasets are needed to validate the generalization of the models. Additionally, exploring other neural network architectures and feature extraction techniques could further improve classification performance. Nevertheless, the results of this research suggest that CNNs, combined with data augmentation techniques, represent a promising tool for the diagnosis of SMS.

Author Contributions: Conceptualization, E.N.-V.; methodology, E.N.-V. and R.F.-R.; software, R.F.-R., E.N.-V. and F.D.-M.; validation, A.Á.-M. and E.G.-H.; formal analysis, A.Á.-M.; investigation, E.N.-V. and I.H.-d.; resources, I.H.-d., E.N.-V. and R.F.-R.; data curation, I.H.-d.; writing—original draft preparation, E.N.-V., R.F.-R. and I.H.-d.; writing—review and editing, E.G.-H. and N.H.-K.; visualization, R.F.-R.; supervision, E.G.-H., C.C. and N.H.-K.; project administration, C.C. and R.M.-O.; funding acquisition, R.M.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Agencia Estatal De Investigación from Spain, grant number PID2023-152984OB-I00 and PID2021-124176OB-I00.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Universidad Politécnica de Madrid in 2018 (20180618_TECA).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author (due to the sensitive biometrics nature of the data and relevant data protection regulations).

Acknowledgments: The authors extend our deepest gratitude to the Spanish Smith–Magenis Association for their unwavering collaboration and dedication to our research study. Their invaluable contributions not only enriched our work but also underscore the importance of community engagement in scientific endeavors. The authors are profoundly thankful for their commitment to advancing knowledge and improving lives within the Smith–Magenis community.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SMS	Smith–Magenis syndrome
CNN	Convolutional neuronal network
ROS	Random oversampling
RUS	Random undersampling
SMOTE	Synthetic minority oversampling technique
ADOMS	Adaptive density oversampling along the first principal component axis
CCR	Combined cleaning and resampling
CE-SMOTE	Cluster ensemble synthetic minority oversampling technique
Cluster-SMOTE	Clustered synthetic minority oversampling technique
DBSMOTE	Density-based synthetic minority oversampling technique
DE	Differential evolution
G-SMOTE	Geometric SMOTE
LVQ SMOTE	Learning vector quantization, synthetic minority oversampling technique
NDO-Sampling	Normal distribution oversampling

NEATER	Noncooperative game theory for filtering oversampled data
ProWSyn	Proximity weighted synthetic oversampling technique
SMOBD	Synthetic minority oversampling based on data density
NearMiss	Edited nearest neighbors
SC	Study case
CPP	Cepstral peak prominence
MFCC	Mel-frequency cepstral coefficients

References

- Gómez-Vilda, P.; Palacios-Alonso, D.; Rodellar-Biarge, V.; Álvarez Marquina, A.; Nieto-Lluis, V.; Martínez-Olalla, R. Parkinson's disease monitoring by biomechanical instability of phonation. *Neurocomputing* **2017**, *255*, 3–16. [\[CrossRef\]](#)
- Gómez-Vilda, P.; Mekyska, J.; Ferrández, J.M.; Palacios-Alonso, D.; Gómez-Rodellar, A.; Rodellar-Biarge, V.; Galaz, Z.; Smekal, Z.; Eliasova, I.; Kostalova, M.; et al. Parkinson disease detection from speech articulation neuromechanics. *Front. Neuroinform.* **2017**, *11*, 56. [\[CrossRef\]](#) [\[PubMed\]](#)
- Górriz, J.M.; Ramírez, J.; Ortíz, A.; Martínez-Murcia, F.J.; Segovia, F.; Suckling, J.; Leming, M.; Zhang, Y.D.; Álvarez-Sánchez, J.R.; Bologna, G.; et al. Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. *Neurocomputing* **2020**, *410*, 237–270. [\[CrossRef\]](#)
- Palo, F.D.; Parde, N. Enriching Neural Models with Targeted Features for Dementia Detection. In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Student Research Workshop, Florence, Italy, 28 July–2 August 2019; pp. 302–308. [\[CrossRef\]](#)
- Pereira, C.R.; Weber, S.A.; Hook, C.; Rosa, G.H.; Papa, J.P. Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2016, Sao Paulo, Brazil, 4–7 October 2016; pp. 340–346. [\[CrossRef\]](#)
- Zhang, S.; Poon, S.K.; Vuong, K.; Sneddon, A.; Loy, C.T. A Deep Learning-Based Approach for Gait Analysis in Huntington Disease. *Stud. Health Technol. Inform.* **2019**, *264*, 477–481. [\[CrossRef\]](#) [\[PubMed\]](#)
- Qiu, Y.; Lin, Q.H.; Kuang, L.D.; Zhao, W.D.; Gong, X.F.; Cong, F.; Calhoun, V.D. Classification of Schizophrenia Patients and Healthy Controls Using ICA of Complex-Valued fMRI Data and Convolutional Neural Networks. In *Lecture Notes in Computer Science, Proceedings of the 16th International Symposium on Neural Networks, Moscow, Russia, 10–12 July 2019*; Springer: Cham, Switzerland, 2019; 11555 LNCS, pp. 540–547. [\[CrossRef\]](#)
- Orimaye, S.; Wong, J.S.M.; Fernandez, J.S.G. Deep-Deep Neural Network Language Models for Predicting Mild Cognitive Impairment. In Proceedings of the Advances in Bioinformatics and Artificial Intelligence, New York, NY, USA, 11 July 2016.
- Karlekar, S.; Niu, T.; Bansal, M. Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models. In Proceedings of the NAACL HLT 2018–2018 Conference of the North American Chapter Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, 1–6 June 2018; Volume 2, pp. 701–707. [\[CrossRef\]](#)
- Dakka, J.; Bashivan, P.; Gheiratmand, M.; Rish, I.; Jha, S.; Greiner, R. Learning Neural Markers of Schizophrenia Disorder Using Recurrent Neural Networks. *arXiv* **2017**, arXiv:1712.00512.
- Sivaranjini, S.; Sujatha, C.M. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed. Tools Appl.* **2020**, *79*, 15467–15479. [\[CrossRef\]](#)
- Vásquez-Correa, J.C.; Arias-Vergara, T.; Orozco-Arroyave, J.R.; Eskofier, B.; Klucken, J.; Nöth, E. Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1618–1630. [\[CrossRef\]](#)
- Srinivasagopalan, S.; Barry, J.; Gurupur, V.; Thankachan, S. A deep learning approach for diagnosing schizophrenic patients. *J. Exp. Theor. Artif. Intell.* **2019**, *31*, 803–816. [\[CrossRef\]](#)
- Oh, S.L.; Hagiwara, Y.; Raghavendra, U.; Yuvaraj, R.; Arunkumar, N.; Murugappan, M.; Acharya, U.R. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput. Appl.* **2020**, *32*, 10927–10933. [\[CrossRef\]](#)
- Ali, L.; Zhu, C.; Zhang, Z.; Liu, Y. Automated Detection of Parkinson's Disease Based on Multiple Types of Sustained Phonations Using Linear Discriminant Analysis and Genetically Optimized Neural Network. *IEEE J. Transl. Eng. Health Med.* **2019**, *7*, 1–10. [\[CrossRef\]](#)
- Guatelli, R.; Aubin, V.I.; Pérez, S.N. Predicción de la enfermedad de Parkinson utilizando redes neuronales convolucionales. In *XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020, El Calafate, Santa Cruz)*; Universidad Nacional de la Patagonia Austral: Río Gallegos, Argentina, 2020.
- Haq, A.U.; Li, J.; Memon, M.H.; Khan, J.; Din, S.U.; Ahad, I.; Sun, R.; Lai, Z. Comparative Analysis of the Classification Performance of Machine Learning Classifiers and Deep Neural Network Classifier for Prediction of Parkinson Disease. In Proceedings of the 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 14–16 December 2018; pp. 101–106. [\[CrossRef\]](#)
- Gürüler, H. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Comput. Appl.* **2017**, *28*, 1657–1666. [\[CrossRef\]](#)
- National Organization for Rare Disorders | NORD. Available online: <https://rarediseases.org/> (accessed on 2 May 2024).
- Montserrat, A.; Taruscio, D. Policies and actions to tackle rare diseases at European level. *Ann. Dell'Istituto Super. Sanità* **2019**, *55*, 296–304.

21. de la Paz, M.; Alonso, V.; Zurriaga, O.; Astray, J.; Aldana-Espinal, J.M.; Margolles, M.J.; Jiménez, J.; Palomar, J.A.; Santana, M.; Ramalle-Gomara, E.; et al. National Rare Disease Registries: Overview from Spain. *Orphanet J. Rare Dis.* **2014**, *9*, O8. [CrossRef]
22. EURORDIS. EURORDIS—Rare Disease Europe. Available online: <https://www.eurordis.org/es/> (accessed on 3 May 2024).
23. Albertini, G.; Bonassi, S.; Dall’Armi, V.; Giachetti, I.; Giaquinto, S.; Mignano, M. Spectral analysis of the voice in Down Syndrome. *Res. Dev. Disabil.* **2010**, *31*, 995–1001. [CrossRef] [PubMed]
24. Moore, J.; Thibeault, S. Insights Into the Role of Elastin in Vocal Fold Health and Disease. *J. Voice* **2012**, *26*, 269–275. [CrossRef]
25. Jeffery, T.; Cunningham, S.; Whiteside, S.P. Analyses of Sustained Vowels in Down Syndrome (DS): A Case Study Using Spectrograms and Perturbation Data to Investigate Voice Quality in Four Adults With DS. *J. Voice* **2018**, *32*, 644.e11–644.e24. [CrossRef]
26. Krishnamurthy, R.; Ramani, S.A. Aerodynamic and acoustic characteristics of voice in children with down syndrome-A systematic review. *Int. J. Pediatr. Otorhinolaryngol.* **2020**, *133*, 109946. [CrossRef]
27. la Guía, I.H.D.; Garayzábal-Heinze, E.; Gómez-Vilda, P.; Martínez-Olalla, R.; Palacios-Alonso, D. Acoustic Analysis of Phonation in Children With Smith–Magenis Syndrome. *Front. Hum. Neurosci.* **2021**, *15*. [CrossRef]
28. Antonell, A.; Campo, M.D.; Flores, R.; Campuzano, V.; Pérez-Jurado, L.A. Síndrome de Williams: Aspectos clínicos y bases moleculares. *Rev. Neurol.* **2006**, *42*, 69–75. [CrossRef]
29. Hidalgo, I.; Vilda, P.G.; Garayzábal, E. Biomechanical Description of Phonation in Children Affected by Williams Syndrome. *J. Voice* **2018**, *32*, 515.e15–515.e28. [CrossRef]
30. la Guía, I.H.D.; Garayzábal-Heinze, E.; Gómez-Vilda, P. Voice Characteristics in Smith–Magenis Syndrome: An Acoustic Study of Laryngeal Biomechanics. *Languages* **2020**, *5*, 31. [CrossRef]
31. la Guía, I.H.D.; Garayzábal, E.; Gómez-Vilda, P.; Palacios-Alonso, D. Specificities of phonation biomechanics in Down Syndrome children. *Biomed. Signal Process. Control* **2021**, *63*, 102219. [CrossRef]
32. Palacios-Alonso, D. MonParLoc: A Speech-Based System for Parkinson’s Disease Analysis and Monitoring. *IEEE Access* **2020**, *8*, 184678–184688. [CrossRef]
33. Zhang, N.; Zang, T. A multi-network integration approach for measuring disease similarity based on ncRNA regulation and heterogeneous information. *BMC Bioinform.* **2022**, *23*, 89. [CrossRef] [PubMed]
34. Síndrome de Smith-Magenis. Available online: <https://rarediseases.info.nih.gov/espanol/11853/sindrome-de-smith-magenis> (accessed on 5 November 2023).
35. Elsea, S.H.; Girirajan, S.S. Smith-Magenis syndrome. *Eur. J. Hum. Genet. EJHG* **2008**, *16*, 412–421. [CrossRef] [PubMed]
36. Finucane, B.; Savatt, J.M.; Shimelis, H.; Girirajan, S.; Myers, S.M. Birt-Hogg-Dubé symptoms in Smith-Magenis syndrome include pediatric-onset pneumothorax. *Am. J. Med. Genet. Part A* **2021**, *185*, 1922–1924. [CrossRef]
37. Edelman, E.A.; Girirajan, S.; Finucane, B.; Patel, P.I.; Lupski, J.R.; Smith, A.C.; Elsea, S.H. Gender, genotype, and phenotype differences in Smith-Magenis syndrome: A meta-analysis of 105 cases. *Clin. Genet.* **2007**, *71*, 540–550. [CrossRef]
38. Vlangos, C.N.; Yim, D.K.; Elsea, S.H. Refinement of the Smith–Magenis syndrome critical region to ~950 kb and assessment of 17p11.2 deletions. Are all deletions created equally? *Mol. Genet. Metab.* **2003**, *79*, 134–141. [CrossRef]
39. Slager, R.E.; Newton, T.L.; Vlangos, C.N.; Finucane, B.; Elsea, S.H. Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat. Genet.* **2003**, *33*, 466–468. [CrossRef]
40. Elsea, S.H.; Girirajan, S. Smith-Magenis Syndrome. In *Encyclopedia of Molecular Mechanisms of Disease*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1947–1948. [CrossRef]
41. Falco, M.; Amabile, S.; Acquaviva, F. RAI1 gene mutations: Mechanisms of Smith–Magenis syndrome. *Appl. Clin. Genet.* **2017**, *10*, 85–94. [CrossRef]
42. Vieira, G.H.; Rodriguez, J.D.; Carmona-Mora, P.; Cao, L.; Gamba, B.F.; Carvalho, D.R.; de Rezende Duarte, A.; Santos, S.R.; De Souza, D.H.; DuPont, B.R.; et al. Detection of classical 17p11. 2 deletions, an atypical deletion and RAI1 alterations in patients with features suggestive of Smith–Magenis syndrome. *Eur. J. Hum. Genet.* **2012**, *20*, 148–154. [CrossRef] [PubMed]
43. Greenberg, F.; Guzzetta, V.; Oca-Luna, R.M.D.; Magenis, R.E.; Smith, A.C.; Richter, S.F.; Kondo, I.; Dobyns, W.B.; Patel, P.I.; Lupski, J.R. Molecular analysis of the Smith-Magenis syndrome: A possible contiguous-gene syndrome associated with del(17)(p11.2). *Am. J. Hum. Genet.* **1991**, *49*, 1207. [PubMed]
44. Smith, A.C.; Gropman, A.L. SMITH-MAGENIS SYNDROME. *Cassidy Allanson’s Manag. Genet. Syndr.* **2021**, *54*, 863–893. [CrossRef]
45. Gropman, A.L.; Duncan, W.C.; Smith, A.C. Neurologic and developmental features of the Smith-Magenis syndrome (del 17p11.2). *Pediatr. Neurol.* **2006**, *34*, 337–350. [CrossRef]
46. Angriman, M.; Caravale, B.; Novelli, L.; Ferri, R.; Bruni, O. Sleep in children with neurodevelopmental disabilities. *Neuropediatrics* **2015**, *46*, 199–210. [CrossRef]
47. Wolters, P.L.; Gropman, A.L.; Martin, S.C.; Smith, M.R.; Hildenbrand, H.L.; Brewer, C.C.; Smith, A.C. Neurodevelopment of Children Under 3 Years of Age With Smith-Magenis Syndrome. *Pediatr. Neurol.* **2009**, *41*, 250–258. [CrossRef]
48. Shayota, B.J.; Elsea, S.H. Behavior and sleep disturbance in Smith-Magenis syndrome. *Curr. Opin. Psychiatry* **2019**, *32*, 73–78. [CrossRef]
49. Greenberg, F.; Lewis, R.A.; Potocki, L.; Glaze, D.; Parke, J.; Killian, J.; Murphy, M.A.; Williamson, D.; Brown, F.; Dutton, R.; et al. Multi-disciplinary clinical study of Smith-Magenis syndrome (deletion 17p11. 2). *Am. J. Med. Genet.* **1996**, *62*, 247–254. [CrossRef]
50. Howlin, P.; Udwin, O. *Outcomes in Neurodevelopmental and Genetic Disorders*; Cambridge University Press: Cambridge, UK, 2002.

51. Martin, S.C.; Wolters, P.L.; Smith, A.C. Adaptive and maladaptive behavior in children with Smith-Magenis Syndrome. *J. Autism Dev. Disord.* **2006**, *36*, 541–552. [\[CrossRef\]](#)
52. Brendal, M.A.; King, K.A.; Zalewski, C.K.; Finucane, B.M.; Introne, W.; Brewer, C.C.; Smith, A.C. Auditory Phenotype of Smith-Magenis Syndrome. *J. Speech Lang. Hear. Res. JSLHR* **2017**, *60*, 1076. [\[CrossRef\]](#)
53. Webber, C. *Cognitive and Behavioural Characteristics of Children with Smith-Magenis Syndrome*; University of Leicester: Leicester, UK, 1999.
54. Gómez-Vilda, P.; Álvarez Marquina, A.; Tsanas, A.; Lázaro-Carrascosa, C.A.; Rodellar-Biarge, V.; Nieto-Lluis, V.; Martínez-Olalla, R. Phonation biomechanics in quantifying parkinson's disease symptom severity. *Smart Innov. Syst. Technol.* **2016**, *48*, 93–102. [\[CrossRef\]](#)
55. Tsanas, A. Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning. Ph.D. Thesis, Oxford University, Oxford, UK, 2012.
56. Gómez-Vilda, P.; Rodellar-Biarge, V.; Nieto-Lluis, V.; Ipiña, K.L.D.; Álvarez Marquina, A.; Martínez-Olalla, R.; Ecay-Torres, M.; Martínez-Lage, P. Phonation biomechanic analysis of Alzheimer's Disease cases. *Neurocomputing* **2015**, *167*, 83–93. [\[CrossRef\]](#)
57. Sapir, S.; Ramig, L.O.; Spielman, J.L.; Fox, C. Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *J. Speech Lang. Hear. Res. JSLHR* **2010**, *53*, 114–125. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Skodda, S.; Visser, W.; Schlegel, U. Vowel articulation in Parkinson's disease. *J. Voice* **2011**, *25*, 467–472. [\[CrossRef\]](#) [\[PubMed\]](#)
59. DuHadway, C.M.; Hustad, K.C. Contributors to Intelligibility in Preschool- Aged Children with Cerebral Palsy. *J. Med. Speech-Lang. Pathol.* **2012**, *20*, 11.
60. Levy, E.S.; Leone, D.; Moya-Galé, G.; Hsu, S.C.; Chen, W.; Ramig, L. Vowel Intelligibility in Children With and Without Dysarthria: An Exploratory Study. *Commun. Disord. Q.* **2016**, *37*, 171–179. [\[CrossRef\]](#)
61. Gómez, P.; Mekyska, J.; Gómez, A.; Palacios, D.; Rodellar, V.; Álvarez, A. Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics. *Biomed. Signal Process. Control* **2019**, *52*, 312–320. [\[CrossRef\]](#)
62. Little, M.A.; McSharry, P.E.; Hunter, E.J.; Spielman, J.; Ramig, L.O. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Bio-Med. Eng.* **2009**, *56*, 1015. [\[CrossRef\]](#)
63. Chen, H.L.; Huang, C.C.; Yu, X.G.; Xu, X.; Sun, X.; Wang, G.; Wang, S.J. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst. Appl.* **2013**, *40*, 263–271. [\[CrossRef\]](#)
64. Bot, B.M.; Suver, C.; Neto, E.C.; Kellen, M.; Klein, A.; Bare, C.; Doerr, M.; Pratap, A.; Wilbanks, J.; Dorsey, E.R.; et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* **2016**, *3*, 160011. [\[CrossRef\]](#)
65. Arora, S.; Tsanas, A. Discrimination of Parkinson's disease participants from healthy controls using telephone-quality voice recordings. *Mov. Disord.* **2016**, *31*, S266.
66. Iyer, A.; Kemp, A.; Rahmatallah, Y.; Pillai, L.; Glover, A.; Prior, F.; Larson-Prior, L.; Virmani, T. A machine learning method to process voice samples for identification of Parkinson's disease. *Sci. Rep.* **2023**, *13*, 20615. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Almeida, J.S.; Filho, P.P.R.; Carneiro, T.; Wei, W.; Damaševičius, R.; Maskeliūnas, R.; de Albuquerque, V.H.C.; Almeida, J.S.; Filho, P.P.R.; Carneiro, T.; et al. Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *PaReL* **2019**, *125*, 55–62. [\[CrossRef\]](#)
68. Gunduz, H. Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets. *IEEE Access* **2019**, *7*, 115540–115551. [\[CrossRef\]](#)
69. Zhao, Y.; He, L. Deep learning in the EEG diagnosis of Alzheimer's disease. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ACCV 2014 Workshops, Singapore, 1–2 November 2014*; Springer: Cham, Switzerland, 2015; Volume 9008, pp. 340–353. [\[CrossRef\]](#)
70. Payan, A.; Montana, G. Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks. *arXiv* **2015**, arXiv:1502.02506.
71. Ebrahimighahnavieh, M.A.; Luo, S.; Chiong, R. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Comput. Methods Programs Biomed.* **2020**, *187*, 105242. [\[CrossRef\]](#)
72. Nguyen, H.; Chu, N.N. An Introduction to Deep Learning Research for Alzheimer's Disease. *IEEE Consum. Electron. Mag.* **2021**, *10*, 72–75. [\[CrossRef\]](#)
73. Lopez-De-Ipina, K.; Martinez-De-Lizarduy, U.; Calvo, P.M.; Beitia, B.; Garcia-Melero, J.; Ecay-Torres, M.; Estanga, A.; Faundez-Zanuy, M. Analysis of Disfluencies for automatic detection of Mild Cognitive Impairment: A deep learning approach. In *Proceedings of the 2017 International Work Conference on Bio-Inspired Intelligence: Intelligent Systems for Biodiversity Conservation, IWobi 2017—Proceedings, Funchal, Portugal, 10–12 July 2017*. [\[CrossRef\]](#)
74. Lee, G.; Nho, K.; Kang, B.; Sohn, K.A.; Kim, D. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* **2019**, *9*, 1952. [\[CrossRef\]](#)
75. Bringas, S.; Salomón, S.; Duque, R.; Lage, C.; Montaña, J.L. Alzheimer's Disease stage identification using deep learning models. *J. Biomed. Inform.* **2020**, *109*, 103514. [\[CrossRef\]](#)
76. Mahajan, P.; Baths, V. Acoustic and Language Based Deep Learning Approaches for Alzheimer's Dementia Detection From Spontaneous Speech. *Front. Aging Neurosci.* **2021**, *13*, 623607. [\[CrossRef\]](#)
77. Vachhani, B.; Bhat, C.; Kopparapu, S.K. Data Augmentation Using Healthy Speech for Dysarthric Speech Recognition. In *Proceedings of the Interspeech, Hyderabad, India, 2–6 September 2018*; pp. 471–475.

78. Liu, S.; Geng, M.; Hu, S.; Xie, X.; Cui, M.; Yu, J.; Liu, X.; Meng, H. Recent Progress in the CUHK Dysarthric Speech Recognition System. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 2267–2281. [\[CrossRef\]](#)
79. Shahamiri, S.R. Speech Vision: An End-to-End Deep Learning-Based Dysarthric Automatic Speech Recognition System. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 852–861. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Joshy, A.A.; Rajan, R. Automated Dysarthria Severity Classification Using Deep Learning Frameworks. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2021; pp. 116–120. [\[CrossRef\]](#)
81. Joshy, A.A.; Rajan, R. Automated Dysarthria Severity Classification: A Study on Acoustic Features and Deep Learning Techniques. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 1147–1157. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Hashan, A.M.; Dmitrievich, C.R.; Valerievich, M.A.; Vasilyevich, D.D.; Alexandrovich, K.N.; Andreevich, B.B. Deep Learning Based Speech Recognition for Hyperkinetic Dysarthria Disorder. In Proceedings of the 2024 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT), Yekaterinburg, Russia, 13–15 May 2024; pp. 12–15. [\[CrossRef\]](#)
83. Shanmugapriya, P.; Mohan, V. Comparative analysis of deep learning models for dysarthric speech detection. *Soft Comput.* **2024**, *28*, 5683–5698. [\[CrossRef\]](#)
84. Riis, P. Thirty years of bioethics: The Helsinki Declaration 1964-2003. *New Rev. Bioeth.* **2003**, *1*, 15–25. [\[CrossRef\]](#)
85. Ahmed, S.F.; Alam, M.S.B.; Hassan, M.; Rozbu, M.R.; Ishtiaq, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A.; Gandomi, A.H. Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artif. Intell. Rev.* **2023**, *56*, 13521–13617. [\[CrossRef\]](#)
86. Sarker, I.H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2021**, *2*, 420. [\[CrossRef\]](#)
87. Abdani, S.R.; Zulkifley, M.A.; Zulkifley, N.H. Undersampling and oversampling strategies for convolutional neural networks classifier. In *Lecture Notes in Electrical Engineering, Proceedings of the 6th International Conference on Electrical, Control and Computer Engineering: InECCE2021, Kuantan, Pahang, Malaysia, 23 August 2022*; Springer: Singapore, 2022; pp. 1129–1137.
88. Basha, S.J.; Madala, S.R.; Vivek, K.; Kumar, E.S.; Ammannamma, T. A review on imbalanced data classification techniques. In Proceedings of the 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), Coimbatore, India, 4–5 March 2022; pp. 1–6.
89. Yan, Y.; Chen, M.; Shyu, M.L.; Chen, S.C. Deep learning for imbalanced multimedia data classification. In Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM), Miami, FL, USA, 14–16 December 2015; pp. 483–488.
90. Kovács, G. An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. *Appl. Soft Comput.* **2019**, *83*, 105662. [\[CrossRef\]](#)
91. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
92. Hayaty, M.; Muthmainah, S.; Ghufra, S.M. Random and synthetic over-sampling approach to resolve data imbalance in classification. *Int. J. Artif. Intell. Res.* **2020**, *4*, 86–94. [\[CrossRef\]](#)
93. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [\[CrossRef\]](#)
94. Ertekin, Ş. Adaptive oversampling for imbalanced data classification. In *Lecture Notes in Electrical Engineering, Proceedings of the Information Sciences and Systems 2013: Proceedings of the 28th International Symposium on Computer and Information Sciences*; Springer: Cham, Switzerland, 2013; pp. 261–269.
95. Zhou, B.; Yang, C.; Guo, H.; Hu, J. A quasi-linear SVM combined with assembled SMOTE for imbalanced data classification. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–7. [\[CrossRef\]](#)
96. Koziarski, M.; Woźniak, M. CCR: A combined cleaning and resampling algorithm for imbalanced data classification. *Int. J. Appl. Math. Comput. Sci.* **2017**, *27*, 727–736. [\[CrossRef\]](#)
97. Li, K.; Zhang, W.; Lu, Q.; Fang, X. An Improved SMOTE Imbalanced Data Classification Method Based on Support Degree. In Proceedings of the 2014 International Conference on Identification, Information and Knowledge in the Internet of Things, Beijing, China, 17–18 October 2014; pp. 34–38. [\[CrossRef\]](#)
98. Cieslak, D.A.; Chawla, N.V.; Striegel, A. Combating imbalance in network intrusion datasets. In Proceedings of the GrC, San Diego, CA, USA, 10–13 September 2006; pp. 732–737.
99. Bunkhumpornpat, C.; Sinapiromsaran, K.; Lursinsap, C. DBSMOTE: Density-based synthetic minority over-sampling technique. *Appl. Intell.* **2012**, *36*, 664–684. [\[CrossRef\]](#)
100. Chen, L.; Cai, Z.; Chen, L.; Gu, Q. A novel differential evolution-clustering hybrid resampling algorithm on imbalanced datasets. In Proceedings of the 2010 Third International Conference on Knowledge Discovery and Data Mining, Phuket, Thailand, 9–10 January 2010; pp. 81–85.
101. Sandhan, T.; Choi, J.Y. Handling imbalanced datasets by partially guided hybrid sampling for pattern recognition. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 1449–1453.
102. Lee, H.; Kim, J.; Kim, S. Gaussian-based SMOTE algorithm for solving skewed class distributions. *Int. J. Fuzzy Log. Intell. Syst.* **2017**, *17*, 229–234. [\[CrossRef\]](#)

103. Lee, J.; Kim, N.R.; Lee, J.H. An over-sampling technique with rejection for imbalanced class learning. In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, Bali, Indonesia, 8–10 January 2015; pp. 1–6.
104. Nakamura, M.; Kajiwar, Y.; Otsuka, A.; Kimura, H. Lvsq-smote-learning vector quantization based synthetic minority over-sampling technique for biomedical data. *BioData Min.* **2013**, *6*, 1–10. [[CrossRef](#)] [[PubMed](#)]
105. Zhang, H.; Wang, Z. A Normal Distribution-Based Over-Sampling Approach to Imbalanced Data Classification. In *Lecture Notes in Computer Science, Proceedings of the Advanced Data Mining and Applications, Beijing, China, 17–19 December 2011*; Tang, J., King, I., Chen, L., Wang, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 83–96.
106. Almogahed, B.; Kakadiaris, I. NEATER: Filtering of Over-sampled Data Using Non-cooperative Game Theory. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 1371–1376. [[CrossRef](#)]
107. Barua, S.; Islam, M.M.; Murase, K. ProWSyn: Proximity weighted synthetic oversampling technique for imbalanced data set learning. In *Lecture Notes in Computer Science, Proceedings of the Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, 14–17 April 2013*; Proceedings, Part II 17; Springer: Berlin/Heidelberg, Germany, 2013; pp. 317–328.
108. Cao, Q.; Wang, S. Applying Over-sampling Technique Based on Data Density and Cost-sensitive SVM to Imbalanced Learning. In Proceedings of the 2011 International Conference on Information Management, Innovation Management and Industrial Engineering, Shenzhen, China, 26–27 November 2011; Volume 2, pp. 543–548. [[CrossRef](#)]
109. Hu, J.; He, X.; Yu, D.J.; Yang, X.B.; Yang, J.Y.; Shen, H.B. A new supervised over-sampling algorithm with application to protein-nucleotide binding residue prediction. *PLoS ONE* **2014**, *9*, e107676. [[CrossRef](#)]
110. Arefeen, M.A.; Nimi, S.T.; Rahman, M.S. Neural Network-Based Undersampling Techniques. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 1111–1120. [[CrossRef](#)]
111. Tanimoto, A.; Yamada, S.; Takenouchi, T.; Sugiyama, M.; Kashima, H. Improving imbalanced classification using near-miss instances. *Expert Syst. Appl.* **2022**, *201*, 117130. [[CrossRef](#)]
112. Abdul, Z.K.; Al-Talabani, A.K. Mel frequency cepstral coefficient and its applications: A review. *IEEE Access* **2022**, *10*, 122136–122158. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.