KAUNAS UNIVERSITY OF TECHNOLOGY

OLUSOLA OLUWAKEMI ABAYOMI-ALLI

SMALL DATA ANALYTICS USING ARTIFICIAL INTELLIGENCE METHODS

Doctoral dissertation
Technological Sciences, Informatics Engineering (T 007)

2024, Kaunas

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Informatics, Department of Software Engineering during the period of 2019–2023. The studies were supported by Research Council of Lithuania.

The doctoral right has been granted to Kaunas University of Technology together with Vilnius Gediminas Technical University.

Scientific supervisor:
Prof. Dr. Robertas DAMAŠEVIČIUS (Kaunas University of Technology, Technological Sciences, Informatics Engineering, T 007).

Edited by: English language editor Dr. Armandas Rumšas (Publishing House *Technologija*), Lithuanian language editor Aurelija Gražina Rukšaitė (Publishing House *Technologija*).

Dissertation Defense Board of the Informatics Engineering Science Field:
Assoc. Prof. Dr. Kristina ŠUTIENĖ (Kaunas University of Technology, Natural Sciences, Informatics, N 009) – **chairperson**;
Prof. Dr. Nikolaj GORANIN (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering, T 007);
Senior Researcher Dr. Gražina KORVEL (Vilnius University, Technological Sciences, Informatics Engineering, T 007);
Dr. Zbigniew MARSZAŁEK (Silesian University of Technology, Poland, Technological Sciences, Informatics Engineering, T 007);
Prof. Dr. Simona RAMANAUSKAITĖ (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering, T 007).

The public defense of the dissertation will be held at 2 p.m. on 17 June 2024 at the public meeting of the Dissertation Defense Board of Informatics Engineering Science Field in the Rectorate Hall at Kaunas University of Technology.

Address: K. Donelaičio 73-402, Kaunas LT-44249, Lithuania.
Tel. no. (+370) 608 28 527; e-mail doktorantura@ktu.lt.

The doctoral dissertation was sent out on 17 May 2024.
The doctoral dissertation is available on the internet at http://ktu.edu at the libraries of Kaunas University of Technology (Gedimino 50, Kaunas, LT-44239, Lithuania) and Vilnius Gediminas Technical University (Saulėtekio av. 14, Vilnius, LT-10223, Lithuania).

KAUNO TECHNOLOGIJOS UNIVERSITETAS

OLUSOLA OLUWAKEMI ABAYOMI-ALLI

MAŽŲJŲ DUOMENŲ ANALIZĖ IR TYRIMAS NAUDOJANT DIRBTINIO INTELEKTO METODUS

Daktaro disertacija
Technologijos mokslai, informatikos inžinerija (T 007)

2024, Kaunas

Disertacija rengta 2019–2023 metais Kauno technologijos universiteto Informatikos fakultete, Programų inžinerijos katedroje. Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Doktorantūros teisė Kauno technologijos universitetui suteikta kartu su Vilniaus Gedimino technikos universitetu.

Mokslinis vadovas:

prof. dr. Robertas DAMAŠEVIČIUS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, T 007).

Redagavo: anglų kalbos redaktorius dr. Armandas Rumšas (leidykla „Technologija"), lietuvių kalbos redaktorė Aurelija Gražina Rukšaitė (leidykla „Technologija").

Informatikos inžinerijos mokslo krypties disertacijos gynimo taryba:

doc. dr. Kristina ŠUTIENĖ (Kauno technologijos universitetas, gamtos mokslai, informatika, N 009) – **pirmininkė**;

prof. dr. Nikolaj GORANIN (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija, T 007);

vyresn. m. d. dr. Gražina KORVEL (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija, T 007);

dr. Zbigniew MARSZAŁEK (Silezijos technologijos universitetas, Lenkija, technologijos mokslai, informatikos inžinerija, T 007);

prof. dr. Simona RAMANAUSKAITĖ (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija, T 007).

Disertacija bus ginama viešame Informatikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje 2024 m. birželio 17 d. 14 val. Kauno technologijos universiteto Rektorato salėje.

Adresas: K. Donelaičio g. 73-402, Kaunas LT-44249, Lietuva.
Tel. (+370) 608 28 527; el. paštas doktorantura@ktu.lt

Disertacija išsiųsta 2024 m. gegužės 17 d.

Su disertacija galima susipažinti interneto svetainėje http://ktu.edu, Kauno technologijos universiteto bibliotekoje (Gedimino g. 50, Kaunas, LT-44239, Lietuva) ir Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva).

CONTENTS

7

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AUC	Area Under Curve
BiLSTM	Bidirectional Long Short-Term Memory Networks
CBB	Cassava Bacterial Blight
CBSD	Cassava Brown Streak Disease
CGM	Cassava Green Mottle
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMD	Cassava Mosaic Disease
CNN	Convolutional Neural Network
COCOA	COswara-COvid-Augmented
DAM	Data Augmentation Methods
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
ECOC	SVM – Error-Correcting Output Coding
GAN	Generative Adversarial Networks
GFCC	Gammatone Frequency Cepstral Coefficients
KNN	K-Nearest Neighbor
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
ML	Machine Learning
NB	Naïve Bayes
NLP	Natural Language Processing
NN	Neural Network
PCA	Principal Component Analysis
PD	Parkinson Disease
RF	Random Forest
RGB	Red-Green-Blue
RL	Reinforcement Learning
RNN	Recurrent Neural Networks,
ROC	Receiver Operating Characteristics Curve
SD	Standard Deviation
SMOTE	Synthetic Minority Oversampling Technique
SOA	State-of-the-Art
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbor Embedding
VDRRE	Voronoi Decomposition-based Random Region Erasing
VLT	Vocal Length Tract
YFP	YouTube Facial Palsy

1. INTRODUCTION

1.1. Relevance of the work

The application of Artificial Intelligence has been a successful, fast-growing, and incredibly popular domain in almost every field of research. This is due to the increasing and significant role in computer vision tasks with an effective, less computation-demanding, economical, rigorous, and accurate outcome [1, 3]. However, one of the main issues predominant in the application of AI methods for any decision-making task is the problem of data availability often referred to as small dataset or limited dataset problems [4, 5]. Small data can be referred to in engineering applications as data with fewer than 50 samples, whereas, for academic research, the amount of data is fewer than 30 instances [6]. Small data can be said to be sufficient if the quality of data is rich in terms of capturing, and if the requirements for diversity, fairness, and transparency are met. On the contrary, most of the available small datasets are not significant enough to build a study or create a hypothesis due to the lack of some imperative information and varieties in data analytics applications, majorly in real-life decision tasks [8]. However, integrating other insights and observations would ultimately develop a solution or models for future brands or businesses [7].

Data analytics could be described as the application of computer systems in analyzing large datasets for decision support. This field of science has played a crucial role in several other scientific areas like pattern recognition, operation research, computational intelligence, machine learning, etc. Recent studies have shown that small data analytics is as important as big data analytics, especially in real-time or informed decision-making tasks. Previous research studies in image processing tasks emphasized that the use of better-quality small, sampled datasets can produce excellent outcomes to a low-quality big data sample [9]. However, in the real-world application of data analytics ranging from disease detection to classification tasks, there is a pressing need for sufficient and balanced datasets as an essential key for effective and efficient performance in the application of AI methods. However, the application of small datasets and artificial intelligence methods, especially in domains such as healthcare, plant disease, sound detection, manufacturing, etc. still suffers from biases, computation complexity, imbalanced datasets, poor generalizability, etc.

Furthermore, small datasets issues in the medical health domain remain a prevailing challenge due to several reasonable factors affecting data collection, such as data privacy based on sensitive/vital patient information (personal, confidential data), cost implication, manpower and expertise requirements for annotation, and anonymized patient's information, before using or sharing this data [10]. Other problems of small datasets in AI applications include a high variance, resulting in skewed outcomes or over-generalization of the data-driven models [4]. These limitations of small data are serious concerns affecting all domains of scientific research, healthcare, industries, etc. Hence, it requires urgent research focus on providing effective models to solve small data problems. Based on the recent success on artificial methods presented in the literature is the application of data augmentation methods which has improved small data problems through the generation of

virtual/synthetic samples using the previous knowledge acquired from a specified limited training data [11]. Data augmentation (DA) application for generating new datasets has shown promising results and significantly improved performance learning classifiers in areas like image recognition, signal and voice recognition, industrial sectors [6], object detection, etc. In other words, one of the most significant methods to efficiently enhance classifier performance on a small data sample is the adoption of a data augmentation approach [12].

Data Augmentation Method (DAM) is a well-known and acceptable pre-processing technique used to expand the variety and quantity of available training data with no direct gathering of additional data samples [13]. The significance of DAM is not limited to merely boosting the variability of the data but also decreasing class imbalance, dataset biases, and preventing overfitting of models. Therefore, it has been effectively applied in image processing and computer vision tasks to synthesize and balance training datasets by creating additional samples of the minority class instances [14, 15]. The impact of data augmentation techniques cuts across other fields where insufficient data has been a major problem such as geothermal field, hardness prediction in *High Entropy Alloy* (HEA), customer credit, etc. This approach is essential for increasing training data, and hence attaining a satisfactory outcome and tackling the issue of insufficient data [16]. Recently, the application of DAM has resulted in the generation of artificial data which has enhanced a crucial part of computer vision research, as the majority of the models that achieved significant outcomes in the *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) [17] were implemented with data augmentation techniques. Examples of the DAMs used in previous literature to improve vision tasks include the following: cropping, flipping images (horizontal and vertical), scaling, color jittering, Gaussian Blur, rotation, noise, etc. [15, 16, 18]. In addition, DAM has proven to be useful in gaining knowledge from skewed datasets. However, in the case of extremely imbalanced classes within a small dataset, applying augmentation methods may not present the expected differences for unique instances in dataset rebalancing [14], but it still contributes to improved performance of classifiers.

Recent studies have shown various data augmentation methods being applied in tackling the challenges of insufficient training data, and, of these, the most popularly used approaches are the conventional or geometric methods such as random rotation, flipping, transformation, etc. Some categories of advanced augmentation techniques proposed in existing literature include *Generative Adversarial Networks* (GANs) and *Variational Autoencoder* (VAE) [19]. The majority of prior studies have indicated progressing results in the application of some data augmentation methods; however, some currently existing approaches still suffer from poor generalization capabilities due to noisy backgrounds, problems with wrong annotations, or label mismatch [20]. In addition, the improvement in the performance of the AI methods is dependent on the size, quality, and variety of the available data [4].

1.2. Object of the Work

The object of this research is to address the challenges of small data analytics by developing and improving data augmentation methods for deep learning model

classification. Different augmentation approaches are presented and applied to various cases of disease detection for both images and sound datasets.

1.3. Aim of the Work

This dissertation aims to improve data augmentation methods for better generalization and modelling of small datasets in image and sound recognition systems.

1.4. Task of the Work

The following tasks have been established to fulfil the intended purpose of the thesis:

1. Review and analyze literature on the existing data augmentation methods to identify efficient methods for the evaluation of small datasets in disease classification.
2. To introduce and apply the developed augmentation methods on benchmark image and sound datasets, by aiming to address small data problems and enhance generalization capabilities and the overall performance.
3. To propose and evaluate the modified augmentation methods for enhancing disease detection, and to apply these methods to images and sound datasets.
4. To investigate and compare the augmentation methods in specific application areas to measure the performance and adaptation to real-life scenarios.

1.5. Practical Value

The practical value of the four tasks and the developed methods outlined in the dissertation is significant in various real-life applications across multiple domains:

1. The novel data augmentation methods used for cassava disease detection have practical applications in agriculture. By employing image augmentation techniques, such as blurring and overexposure, the model can identify plant diseases under varied field conditions more accurately. This approach can lead to more effective plant disease management, thus potentially increasing the crop yields and benefiting farmers and economies reliant on agricultural production. Moreover, in the future, this method could be adapted for other applications, such as industrial quality control. In manufacturing, such as in the automotive or electronics industries, this method could be used to detect defects in products under varied lighting and environmental conditions, thereby enhancing the accuracy of automated inspection systems.
2. The VDRRE image augmentation method for facial palsy detection has important implications in healthcare. The ability to accurately detect and classify facial palsy from images can streamline diagnosis, thus allowing for faster and more effective treatment. This method could be particularly valuable in remote or under-resourced areas where access to specialists is limited. Moreover, in the future, the method could be applied in the development of advanced face image-based emotion recognition systems to better understand human emotions and reactions in real time.

16

3. The covSMOTE augmentation technique for skin melanoma detection addresses a critical need in dermatology and oncology. Data enhancement for the SqueezeNet model improves the detection and classification of skin melanoma, which is crucial for early diagnosis and treatment. This method could be integrated into telemedicine platforms, aiding in the screening process and potentially reducing the burden on healthcare systems. Moreover, in the future, the covSMOTE technique, which focuses on handling class imbalances, could also be adopted for other applications, such as agricultural pest detection. By training models to recognize minor yet critical variations in pest appearances, this method can help in the early detection and management of crop infestations, thereby potentially preventing large-scale agricultural losses.

4. The augmentation method involving noise injection and color transformation for COVID-19 detection has significant implications for the management of the pandemic and public health. The ability to detect COVID-19 from breath sounds by using AI could lead to the development of non-invasive, rapid testing methods. This can facilitate widespread screening, especially in areas where the traditional testing methods are logistically challenging. Moreover, in the future, this method could also be repurposed for other domains of application, such as environmental monitoring, specifically, in detecting and analyzing sound pollution in natural environments or animal sound datasets, thus contributing to biodiversity research and environment conservation efforts.

These methods demonstrate how advanced data augmentation techniques can significantly enhance AI model performance in diverse fields, ranging from agriculture to public health. By improving the accuracy and efficiency in detection and diagnosis, these methods have the potential to positively impact various sectors and contribute to better societal outcomes. By repurposing and adapting these advanced techniques, they could solve a wide array of complex problems in various unrelated domains, thus showcasing the transformative potential of AI and data augmentation in a multitude of sectors.

1.6. Thesis Statements

The application of data augmentation methods has been demonstrated to be a crucial tactic in augmenting the capabilities of artificial intelligence models in several fields. By altering and increasing the training dataset, this method improves its size, diversity, and resilience, which, in turn, boosts the effectiveness of machine learning models in handling challenging tasks.

1. *Voronoi Decomposition Random Region Erasing* (VDRRE) applies random irregular occlusion, by capturing a diverse array of facial expressions and unique variations found in individuals with face palsy. Additionally, the combination of Image Color Histogram Equalization and Image Quality Reduction techniques forms a hybrid augmentation method which introduces diversities in the visual representation of cassava leaf disease images. This simulation includes distinct lighting conditions, image qualities, and levels of noise for the cassava disease recognition task. Furthermore, the application of

the CovSMOTE method facilitates the generation of synthetic samples for the minority class (skin melanoma) while considering the covariance structure of the data. These augmentation techniques, by generating larger and more varied sample collections, effectively improve the deep learning model's generalization and precise classification of unseen images.

2. In the field of sound signal processing, the conversion of sound signals to spectrogram images allows for the application of image augmentation techniques to increase the diversity of training data samples. A more thorough and varied dataset is produced by using techniques including pitch shifting, time stretching, adding noise, and changing spectrogram properties. This enhanced dataset improves the model's training efficacy, thus enabling it to identify and categorize audio signals more accurately in various settings and conditions.

1.7. Scientific Approval

All of the findings in the thesis are unique and are represented by nine publications in total. In the fields of informatics, machine learning, and electronics, there are six international scientific journal articles referred to as *ISI Web of Science* with a citation index and three conference proceedings publications. The remainder is made up of experimental arrangements, brief unpublished observations, or well-known facts.

List of published papers on the dissertation subject:

1. Data Augmentation and Deep Learning Methods in Sound Classification: A Systematic Review. *Electronics, 11*(22), 3795.
2. Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques. *Electronics, 11*(16), 2520.
3. An Ensemble Learning Model for COVID-19 Detection from Blood Test Samples. *Sensors, 22*(6), 2224.
4. Cassava Disease Recognition from Low-Quality Images Using Enhanced Data Augmentation Models and Deep Learning. *Expert Systems, 38*(7), e12746.
5. Few-Shot Learning with a novel Voronoi Tessellation-Based Image Augmentation Method for Facial Palsy Detection. *Electronics, 10*(8), 978.
6. Malignant Skin Melanoma Detection Using Image Augmentation by Oversampling in Nonlinear Lower-Dimensional Embedding Manifolds. *Turkish Journal of Electrical Engineering and Computer Sciences, 29*(8), 2600–2614.
7. BiLSTM with data augmentation using interpolation methods to improve early detection of Parkinson's disease. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*. pp. 371–380. IEEE.
8. Data augmentation using principal component resampling for image recognition by deep learning. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020*, Zakopane, Poland, October 12–

14, 2020, Proceedings, Part II 19 (pp. 39–48). Springer International Publishing.

9. Abayomi-Alli, O. O., Sidekerskienė, T., Damaševičius, R., Siłka, J., & Połap, D. (2020). Empirical Mode Decomposition Based Data Augmentation for Time Series Prediction Using NARX Network. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020*, Zakopane, Poland, October 12–14, 2020, Proceedings, Part I 19 (pp. 702–711). Springer International Publishing.

The full list of publications can be found in the chapter titled LIST OF PUBLICATIONS OF OLUSOLA OLUWAKEMI ABAYOMI-ALLI ON DISSERTATION TOPICS

1.8. Scientific Novelty

The scientific novelty of the dissertation lies in its potential to advance the domain of artificial intelligence by addressing the challenges of small data analysis and improving the outcome of deep learning models through innovative data augmentation approaches. This research presents several novelties of augmentation methods for various unique image data with decision analysis emphasis on the image features for the identified problems.

First, a novel augmentation method based on random erasing has been proposed as the Voronoi Decomposition-based Random Region Erasing approach was adopted for generating variations of synthetic datasets. The detection accuracy showed a better performance of 99.34% accuracy than the currently existing augmentation methods with GAN with an increasing accuracy rate of +4.53%.

Secondly, the convolution of the Chebyshev orthogonal functions with the *Probability Distribution Functions* (PDFs) and the image quality reduction augmentation method has been applied to enhance cassava disease detection. The investigation showed that low-quality images achieved 97.7% accuracy while a +2.0% accuracy increment was obtained with the original (high-quality) and augmented dataset (low-quality) images with an overall accuracy of 99.7%. The proposed augmentation methods showed improved performance in comparison to some of the existing methods using the analyzed datasets.

Third, the investigation of the proposed CovSMOTE augmentation method for skin melanoma detection achieved the optimal detection accuracy of 92.18% improvement in comparison with some of the currently existing methods.

Fourth, this research investigated the influence of data augmentation methods which are color/photometric and noise injection methods in sound datasets. These augmentations have been applied to two feature extraction approaches, specifically, Mel spectrogram and GFCC, and the detection accuracy showed an improvement in the detection rate in comparison to the state-of-the-art methods with an accuracy of 90.1%.

1.9. Thesis Organization

There are six chapters in the thesis, as depicted in Fig. 1.1: The first chapter serves as an introduction and gives a brief overview of the originality, goals, and

objectives of the study. Chapter 2 provides a comprehensive overview of artificial intelligence methods. Small data analytics and data augmentation methods have been analyzed in detail. This chapter further describes in detail the literature review and the different applications domain of data augmentation in small data analysis to get the readers familiarized with terminologies and application areas that will be implemented in the subsequent chapters. Chapter 3 describes the summary of the different publicly accessible small datasets used in this dissertation. In Chapter 4, the materials and the proposed methodology are described. Further architectures and theoretical backgrounds of the proposed augmentation techniques are presented.

Chapter 5 describes the obtained experimental results and explains in detail the comparison results of the proposed methods with state-of-the-art methods. Finally, Chapter 6 summarizes the conclusions and provides future recommendations.

CHAPTER 2

2.1 Overview of Artificial Intelligence

2.2 Overview of Small Data Analytics

2.3 Overview of Data Augmentation Methods (DAM)

2.4 Related Study in Computer Vision Task

2.5 Related Studies- Sound/Audio application

2.6 Gaps in Literature and Research Opportunities

2.7 Review Summary

Literature Review of Artificial Intelligence Methods in Small Data.

CHAPTER 3

3.1 Image Classification Databases

3.2 Sound Classification Databases

3.3 Summary of Dataset used

DATASETS USED FOR THE RESEARCH

CHAPTER 4

4.1 Overview of Research Methodology

4.2 Motivation

4.3 Justification of Proposed Methods

4.4 Proposed Image Augmentation Methods

4.5 Proposed Sound Augmentation Methods

4.6 Performance Metrics

4.7 Summary of the proposed materials and methods

MATERIALS AND METHODOLOGY

CHAPTER 5

5.1 Analysis of face palsy detection

5.2 Analysis of Cassava Disease Detection

5.3 Analysis of Skin Melanoma Classification

5.4 Analysis of COVID-19 detection (Sound dataset)

5.5 Limitation of Methods

5.6 Experiments Summary

EXPERIMENT AND RESULTS

CHAPTER 6

6.1 General Conclusions

6.2 Future Recommendations

GENERAL CONCLUSIONS AND FUTURE RECOMMENDATION

Fig. 1. 1 Dissertation structure and arrangement

21

2. LITERATURE REVIEW OF ARTIFICIAL INTELLIGENCE METHODS FOR SMALL DATA ANALYSIS

In this chapter, analysis of the existing literature is carried out, and a detailed overview of the basic principles of small data analytics is provided. The application of data augmentation methods and machine learning methods for images, and sound data is analyzed, and the performance methods are elaborated upon.

The subsections in this chapter include (2.1) an overview of the artificial intelligence methods, (2.2) an overview of small data, and (2.3) an overview of data augmentation methods. (2.4) serves as a related study in the computer vision task, whereas (2.5) describes the related study in sound classification. (2.6) provides the shortcomings and research gaps in the currently existing studies, and, finally, (2.7) provides the review summary.

2.1. Overview of Artificial Intelligence

Machine Learning (ML) can be referred to as the field of study which enables computers with learning potentials with no specific program. Over the previous decades, a lot of progress has been made in advancing machine learning algorithms by increasing computational power and the volume of data [21]. Future prediction on the relevance of ML algorithm states that a combination of imaging data sets with an increasing growth in computer vision will rapidly increase the overall performance, thus exceeding human accuracy [22]. The future power of clinical applications or medical informatics lies in the effective and efficient application of AI methods in recognizing various diseases [23, 24].

The application of conventional ML approaches in disease detection, such as Support Vector Machines [25], Neural Networks, Naive Bayes, Random Forests, Decision Trees, etc., has been used in categorizing or classifying visuals. Recently, the paradigm shifts in algorithms based on the required data and processing capacity for learning and relearning millions of parameters have attracted focus on research majorly in the deployment of deep learning models for medical image diagnosis [26]. The function and impact of advancement in deep learning and machine learning algorithms for automating medical diagnoses has thereby enhanced the clinical decision-making process [21, 27]. Nevertheless, because of recognizing the paucity of clinical data and the effect of this deficiency on the performance of these learning algorithms, the research focus has been tailored towards maximizing the available small data and increasing the overall model performance [5, 28].

Deep Learning (DL) is a typical AI method which is based on sets of machine learning approaches biologically motivated by the brain structure [29]. The usage of AI methods varies, and the different AI methods include *Convolutional Neural Networks* (CNN), *Recurrent Neural Networks* (RNN), and Reinforcement learning. The adoption of CNN models has shown significant positive results and great consistency, especially in image classification tasks [30]. However, as promising as the outcome of deep learning models mainly in medical imaging systems, it still suffers from inadequate access to larger and well-annotated datasets [21]. Some of the reasons for data scarcity in the medical domain include patient privacy, ethical and legal issues, etc. In the past years, the application of deep learning has been effectively

used in diverse research fields, ranging from computer vision tasks, gaming, speech/audio recognition, natural language processing, etc. [31]. DL methods are the most often used techniques for feature recognition and estimation, especially in image classification tasks [32]. Recently, the advancement in image and object recognition systems has increased the need for improvement in terms of the use of deep learning, such as CNN architectures, in many recognition problems [26, 29]. Subsequently, there have been a series of accomplishments in the adoption of deep learning models in the areas of image detection, object detection, video understanding, etc. [33]. This advancement in the deployment of deep learning models has improved the detection and prediction accuracy, simplified resource workflow, while also providing prowess to underserved populations, and produced biological discoveries and insights [21]. Other deep learning methods presented in the literature include the following: the unsupervised Convolutional Deep Belief Network [34], and the back-propagation neural network [35]. Applying deep learning methods on small datasets is rather challenging [36]. Hence, this study is aimed at presenting an extensive overview of small data analytics, the possibilities, and difficulties to be encountered, and the impacts of AI methodologies in enhancing the prediction and decision-making policy.

One of the significant challenges involved in using learning classifiers for classification and prediction tasks is the issue of small data samples and highly imbalanced classes, which results in poor generalization of training models and the overfitting of such models, etc. [37]. The impact of insufficient data samples in learning classifiers cannot be over-emphasized. Related studies have shown that a larger training dataset enhances classifier generalization characteristics of data, thus reducing the problems of overfitting the training data. The influence of recent studies in the use of learning algorithms, first and foremost regarding the applications of neural networks in identifying patterns in small datasets, is of top interest and relevance. It is also fascinating to know that the application areas of some of these advanced AI approaches still suffer from a lack of adequate or large amounts of real-life data for training these models effectively [38].

2.2. Overview of Small Data Analytics

Small data are defined by their typically low volume, discontinuous gathering, and restricted variety, and are typically created to address specific concerns [39]. The majority of big data is unstructured and ambiguous, thus rendering it undesirable for use in the construction of an efficient deep-learning model. However, bridging the gap between data size balancing and meticulous data processing is vital to developing high-performance deep-learning algorithms [21]. The size of sample data can be said to be small if the number of training samples is relatively slow according to the *Vapnik-Chervonenkis* (VC) dimension [40]. The studies of limited data will continue to thrive irrespective of the rapid expansion of big data because of their track records in providing answers to specific concerns. To maximize the value of small data through combination and sharing, there will be a growing effort to harmonize small data about data standard formats, metadata, and documentation. To do this, the adoption of new data structures will be increasingly used to aggregate, link, and extend the data from these investigations. While, by exposing individuals to the new data science epistemologies, scaling tiny data runs the risk of drawing them into

undesirable activities, such as dataveillance, social sorting, control creep, anticipatory governance, etc. [39]. A comparison table showing the differences between small and big data is illustrated in Tbl. 2.1.

Table 2. 1. Comparison of data types: small data versus big data

Criterion	Small Data	Big Data
Volume	Manageable in size, ranging from bytes to gigabytes. Easier to store and process with traditional tools.	Characterized by massive volumes, often terabytes to petabytes, requiring advanced data processing tools and storage solutions.
Quality	Generally high quality due to the manageability of the dataset. Easier to clean, validate, and ensure accuracy.	Potential quality issues due to the vast size. Ensuring consistency and accuracy becomes more challenging, which leads to possible data noise and errors.
Structure	Typically, well-structured and formatted, making it straightforward to process with standard data analysis tools.	Includes a mix of structured, semi-structured, and unstructured data, necessitating more complex processing methods and technologies.
Velocity	Accumulates at a slower pace, allowing for easier capture and storage without the need for real-time processing.	Characterized by high velocity, with data often streaming in real time. Requires robust infrastructure for rapid processing and analysis.
Exhaustivity	Focuses on specific attributes or characteristics, representing a subset of information.	Aims to be exhaustive, capturing all possible data points related to a particular domain, which leads to a more comprehensive dataset.
Variety	Limited in variety, often confined to specific formats and data types.	Encompasses a wide range of data types, including text, images, videos, sensor data, etc., from a variety of sources.
Scalability	Scalability is often not a primary concern, as the traditional data storage and processing systems are typically sufficient.	Requires scalable architectures and cloud-based solutions to accommodate the rapid growth in data volume and complexity.
Purpose	Suited for targeted, specific applications where high quality and precision are crucial, like in medical research or small-scale market analysis.	Ideal for broader applications where the goal is to derive patterns, trends, and insights from large datasets, such as consumer behavior analysis or large-scale environmental monitoring.

Based on the comparison criterion in Table 2.1, the scope of this work define small data as relatively insufficient in size (volume), characterized by a limited sample (variety) and restricted diversity (exhaustivity). Small dataset-related concerns are regarded as the main key problems in early prediction as these issues are found in some medical records, especially in uncommon genetic diseases such as spinocerebellar ataxia, with extremely few records globally [41]. In this dissertation, the experiment and analysis were done on three data type representations which are images, audio, and numerical data.

Data analytics is the application of powerful algorithms and process computations used to scale up the existing statistical methods, thereby building predictive, simulation and optimization models [42, 43]. Studies in the most recent

decades have shown that the prediction of patterns in small datasets requires future research interest, especially in biological data analysis [44]. The challenges of a small sample size have been stated to be a major factor affecting the inaccuracies of prediction models [45]. The use of huge amounts of data is needed for artificial intelligence training [46]; however, the lack of a huge training dataset is a serious challenge which limits the use of profound DL techniques and results in model over-fitting drastically on small training datasets [29]. Analyzing small data samples is very challenging in image or face recognition systems. This is because subjects with few samples under trained models might show unsatisfying performance mostly when applying single-sample algorithms [47]. In addition, analyzing data, especially in image processing (such as for disease detection), requires continuous and updatable disease labels and models [46]. Thus, learning from small samples will continue to remain a problem in machine learning [38]. The ability to remove vital characteristics from small data is the major focus of small dataset learning. However, learning from small data suffers from a variety of biases, such as high costs of a collection or annotation of suitable data [48], and unclear or prohibited license of data usage [49]. The necessity to create credible models and to secure good inference from a small dataset for an increasing accuracy is still a significant issue [50]. Some of the methods explored by researchers in enhancing the performance of models for the training of small datasets include implementing improved models developed from other huge public datasets, such as *ImageNet* and using artificial training data samples [29, 51].

2.3. Overview of Data Augmentation Methods (DAM)

Data Augmentation is a pre-processing step in machine learning and deep learning where the original dataset is transformed and expanded to create new training samples [52]. This application is popularly used on small datasets to enhance the quantity of the training data and thus overcome the problem of overfitting. Hernández-García et al. [53] summarized the aim of data augmentation as the ability to achieve generalization without reducing the representational strength of the models and changing other hyperparameters. Thus, data augmentation methods can be applied differently, such as linear or non-linear transformation, the addition of auxiliary variables, and data creation using generative models [54]. Gómez-Ríos et al. [55] also described data augmentation as systematically applying deformations to raw input data to significantly increase the quantity of training datasets. The spatial patterns of the target classes should not be altered by the applied deformations. The application of data augmentation cuts across different areas, such as NLP [56, 57], speech recognition [58], facial recognition [59], motion detection [21], disease detection [60], etc. The data augmentation is focused on enhancing the training dataset's volume and diversification, reducing overfitting and improving the robustness and generalization of the model. The augmented data is used to train the ML models, providing additional training examples, and helping to enhance the accuracy of the model. Furthermore, it also boosts the overall efficiency of deep learning models, as well as the general stability, and the standard of training outcomes [61], which serves for solving 'small data problems' [28]. To maximize the reliability of deep learning models on small data, further practical solutions, such as dropout regularization, transfer learning, pretraining, and batch normalization, have been created [56, 62, 63]. Image

augmentation can be achieved through various conventional or random transformations methods, namely, scaling, rotation, flipping [64], width shift, height shift, random cropping [65], shear, zoom, random erasing [66], and dropout regularization [18, 67], addition of noise (salt and pepper, Gaussian noise, Poisson noise), and sample synthesis (using generative models), elastic deformation [68], parameter expansion [69]. Wong et al. [68] applied a normalized random displacement field $u(x, y)$ for every single pixel location (x, y) where a unit displacement vector in an image is specified as $R_w = R_o + \alpha\mu$, where R_w and R_o specify where every pixel is located in the original and warped images. Other interesting approaches are the block-out method [70], cut-out [71], the *Unsharp Masking Method* (USM) [67], and SMOTE [72].

In the recent times, the use of DAM for speech recognition has been an active research area which focuses on using varying procedures to change sound. Examples of methods proposed in previous studies are *Vocal Tract Length Normalisation* (VTLN), Speed perturbation, SpecAugment [58], and Data Warping [68]. Another interesting area where DAM has been deployed is also in reinforcement learning. Jang et al. [73] introduced a K-mixup using the Koopman invariant subspace to increase the training sample size. The duration-modification approach was proposed by [74] in augmenting the training data for the detection of dysarthria speakers. Other methods, such as the addition of background noise, room impulse response, pitch and time shift, etc., have also improved the diversity and variations of training data in sound/audio classification tasks. There are some advanced augmentation methods, such as deep Conv GAN [75], and Wasserstein GAN [76] as well. The outstanding capabilities of DAM methods in sound or audio classification tasks have aided classification performance and the overall competence in many other areas of application. However, the use of DAM techniques in small datasets suffers from low variance within severity classification, thus affecting the performance of the learning model. A categorization of some of the existing DAM techniques in computer vision applications and audio/sound applications is presented in Fig. 2.1, while Tbl. 2.2 shows a summary of categories of augmentation methods and the prevailing limitations. In the following subsections, previous work in the application of AI methods in small data analytics shall be addressed.

Table 2. 2. Comparison of Data Augmentation Methods

References	Categories of Augmentation Methods	Shortcomings of Existing Methods
[77] [78] [50]	Transformation methods (translation, rotation, flipping, scaling, cropping, and shearing)	o Inability to increase generalization capabilities o Generation of highly correlated samples
[71]	Elastic transformation methods	o Creating highly unrealistic samples

[65] [79] [80]	Pixel-level transformations (Photometric methods)	o Possibility of harming the performance of the model when improperly applied
[81] [66] [82]	Random occlusion methods (cutout, cut mix, hide and seek, multicut, etc.)	o The model may result in information loss
[83] [84] [85] [86] [87]Click or tap here to enter text.	Generative (Adversarial) Networks (GANs, VAE)	o The model is prone to mode collapse due to lack of diversity o GANs are extremely difficult to train o The generated samples do not reflect the visual characteristics or distribution of true samples o The model requires high computational cost and complexity (Memory and Time) to generate high-resolution and realistic images
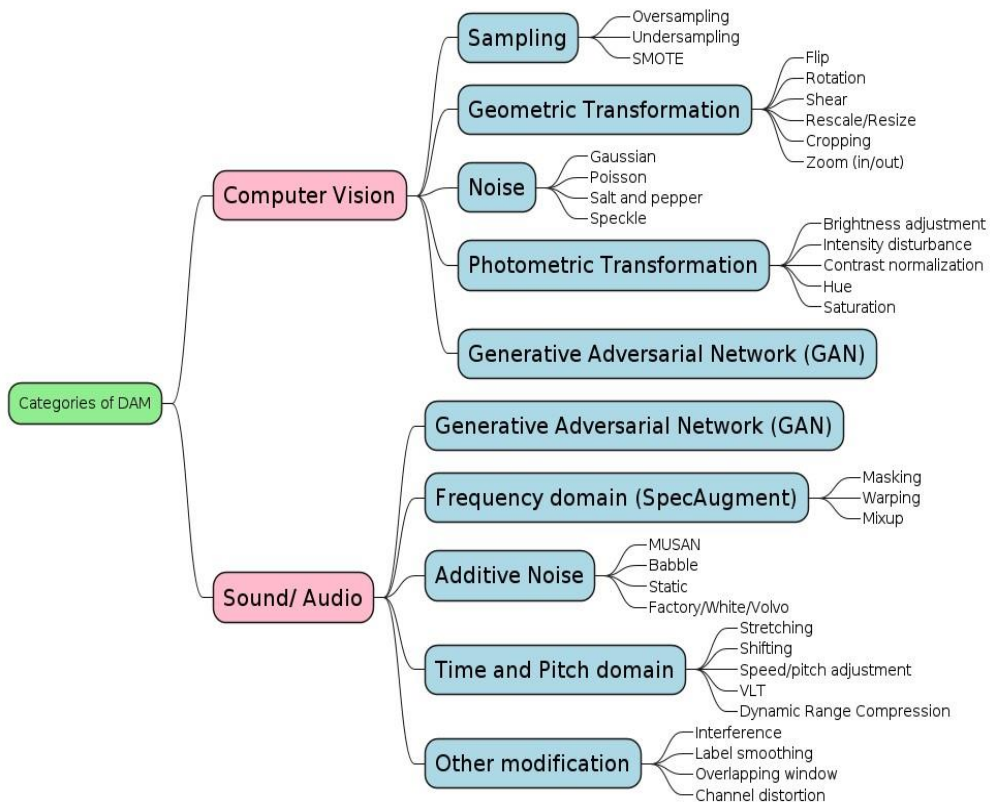
Fig. 2. 1. Categorization of data augmentation methods in the literature. The category is divided into two parts for images and sound classification methods

2.4. Related Study in a Computer Vision Task

In this subsection, a literature review of methods and techniques used in the computer vision domain, especially in application areas like facial palsy, plant disease, and skin melanoma classification, shall be presented [60, 88, 89].

2.4.1. Related work in facial palsy detection

Facial palsy, often known as Bell's palsy, is a severe form of paralysis of the human facial nerves which impairs the ability of the affected facial muscles to contract [90], as demonstrated in Fig. 2.2. On the damaged side of the face, some of the symptoms include a deformed face and dysfunctional facial expressions. Patients with face palsy who are impacted by the disease may have severe disruptions in their daily lives. As the disease advances, there are frequently physical, psychological, and social difficulties that are related to it that can affect a person's quality of life, cause sadness, or cause them to be socially stigmatized [91]. Currently, only experienced doctors can identify facial palsy through an observation of facial symmetry and analysis of dysfunctional facial expressions. The lack of adequate methods for the accurate assessment of the function of the nerves in the facial region, which could be significant in understanding the progression of the disease, is the main obstacle in the detection of facial palsy [92].

Fig. 2. 2. Sample of a palsy face image [89]

While many previous methods used handcrafted features and classifiers, an advanced computer-vison approach for diagnosing and automatically identifying palsy in the face has been developed recently [93]. The electronic facial paralysis assessment method was established in a study by [94], which uses facial motion analysis to examine patients with synkinesis as well as Bell's palsy (involuntary contraction of muscles). To distinguish between tendencies in both local and global regions of facial palsy, authors in [95] suggested multiresolution local binary patterns (LBPs) by measuring and evaluating the resistor-average distance between the facial features and the symmetry of facial movements. The facial palsy symptoms were

measured and evaluated by using SVM. They used a dataset of 197 videos, and their suggested model had an impressive accuracy rate of 94%. An *Active Shape Model* (ASM) was presented by [96] for the recognition of unique features in patients' faces. *Limited-Orientation Modified Circular Gabor Filters* (LO-MCGFs) were suggested by the authors of [97] for use on facial images of 75 palsy faces and 10 healthy faces. Jiang et al. [98] uses laser speckle contrast imaging, while [99] presented a deep hierarchical network (DHN) using cutting-edge *YOLO-v3* [100] architecture. Guo et al. [101] developed a CNN model of prediction and feature extraction by using the evaluation of unilateral peripheral facial paralysis. In Storey et al. [102], the CNN framework was applied, while Meta-learning methods were proposed by [103] to train neural networks to identify new classes, thus addressing the few-shot learning challenge. In addition, the effect of DAN has facilitated the ability of classifiers to learn how to categorize objects on a scale smaller than the norm. Thus, the model can detect different image kinds in various regions of the frame.

The application of data augmentation methods has enhanced the neural network capacity for generalization while preventing overfitting. Some of the data augmentation methods applied in face palsy detection include the conventional methods, such as flipping, cropping, masking, rotations, GAN, etc. Meanwhile, Ten-Harkel et al. [104] concluded that the introduction of geometric transformation methods only gave a relatively small improvement when using augmentation methods in facial palsy detection. The impact of data augmentation methods achieved a mean accuracy of $98 \pm 1\%$ and $90 \pm 7\%$ without augmentation [105], while AUC performance with DAM methods scored $76 \pm 0.04\%$ [106], yet the best accuracy achieved by using the traditional DAM gave 67.24% in comparison to the previous methods [107]. Some of the shortcomings of the existing DAM methods in facial palsy detection include the following: poor quality of augmented images, lack of diversity in the intensity levels of palsy, a very noisy dataset with the least correlation or relationships between the original and the synthetic dataset, a low specificity rate, etc.

2.4.2. Related work in cassava disease classification

In the sub-Saharan region of Africa, cassava is, in fact, the main significant crop which serves as a source of carbohydrates for human sustenance and is cultivated by almost 80% of small farmer households [108, 109]. This crop has demonstrated its value in enhancing the African economy by boosting exports, creating jobs, and reducing poverty [110]. Considering the need to ensure zero hunger (sustainable development goals) to meet the growing global population, cassava yields are still inadequate for sustenance. Some of the factors with a detrimental impact on crop yield and productivity include out-of-control diseases constantly affecting plant leaves (see Fig. 2.3), soil fertility issues, and erratic climate change patterns [111, 112].

Fig. 2. 3. Sample of cassava diseased leaf (left); healthy leaf (right)

To promote and enhance early diagnosis and detection of cassava infections and consequently increase crop yields in general, machine learning techniques must be applied. However, there is still a major problem with getting sufficient data for machine learning algorithms to recognize plant diseases. Secondly, time complexity and the high cost of hiring specialists to annotate these diseases need to be overcome [113]. The importance of image quality has been given minimal attention in many computer vision models. Although most AI models for image recognition tasks have been trained on datasets of better-resolution images, the input images for real-world applications cannot be assumed to be of high quality [114, 115]. In addition, while prior work on the use of NN produced some good outcomes in large part due to the use of better-quality datasets, actual applications suffer from low-quality images that are impacted by a variety of issues, such as a limited dynamic range, noise, and poor resolution, etc. [116]. Hence, the classification of poor-quality images, or, in the worst case, synthetic samples is a challenging and intriguing study [117, 118].

The introduction of the data augmentation approach has helped significantly to promote effective classification as many of the existing augmentation methods proposed by prior studies have been elaborated mainly on an increase of the training samples. Interesting cases of augmentation – such as geometric transformation (zoom, shifting, shear, translation, reflection, random rotations, flipping, etc.) have been safely employed. These simple augmentation methods have proven effective as the original feature representation of the data is not lost during transformation. Other advanced DAM methods, such as variants of GAN, autoencoder variation method, etc., have been used to improve the model generalization. Incorporation of several geometric methods has also improved the training of learning models by boosting the model's ability to learn different variations of data samples and thus assisting in mitigating the overfitting.

In addition, in order to promote and enhance early diagnosis and detection of cassava illnesses and increase crop yields, effective machine learning methods must be applied. The difficulty of gathering adequate data for machine learning programs to recognize plant diseases is still a significant issue. DAM methods, such as AR-

30

GAN, obtained an increased classification outcome of +5.2%, LeafGAN improved the detection classification by +7.4% [119], CycleGAN achieved an F1-score of 81.6% [120], DCGAN and WGAN methods achieved a performance increment of +17% F1-score rate [121], modified CycleGAN achieved the best accuracy of 67.1 ± 0.010 % in comparison with the baseline 62.6 ± 0.015 % [122].

According to recent studies, deep learning-based approaches have consistently shown outstanding outcomes in diverse study fields [123–125]. The CNN adoption is more effective in solving image classification tasks due to its superior capacity to extract useful attributes from data [126] in comparison to the traditional ML algorithms [127]. Some interesting studies in plant disease classification methods include the weak DenseNet-16 architecture [128], a classification using two deep learning architectures GoogleNet and AlexNet [124], Deep Residual Neural Network [129–131], ResNet and SegNet models [132], VGG-16 and 19 [133], a depthwise separable CNN based on MobileNet and VGG [134], VGG16, ResNet, Inception V4, and DenseNets [135], adaptive AANN [136], and the radial basis function neural network [137, 138].

2.4.3. Related work in skin cancer detection

Nearly 100,000 new occurrences of malignant melanoma are recorded every year, making it one of the deadliest forms of skin cancer in Europe [139]. Estimates reported by the World Health Organization (WHO) stated that 12 million extra people will be at risk of passing away by the end of 2030, accounting for nearly 13% of the worldwide mortality rate [140]. According to recent research, the most important predictive factor for survival in this disease is early identification [141]. Now, hospital pathologists rely on ocular examinations for a large portion of the patients they review. Most of the time, in typical clinical settings, dermoscopic melanoma detection sensitivity is less than 80% [142]. It might be quite difficult for a pathologist who lacks the expertise to differentiate melanoma from nonmelanoma tumors. The manual approach to finding skin lesions requires a significant human effort and is error-prone given the level of knowledge and the amount of time needed for visual inspection [143]. The use of computer-aided processes needs to be improved to accurately detect or minimize the risk of delayed detection of this cancer, thus enhancing the survival rates [144]. Due to apparent similarities, intraclass differences, and artefacts, extracting the relevant features from dermoscopic images by exploiting manually created texture and color features may be exceedingly difficult [145]. As a result, real-time dermoscopic image analysis is still a popular research area for artificial intelligence researchers. The development of AI techniques which would benefit patients and physicians by intelligently identifying and predicting skin cancer disorders is the focus of the current research efforts. Considering the issue with the patient privacy leading to limited skin melanoma publicly available datasets, data augmentation has helped to pave the way to addressing the issue of data scarcity by applying augmentation methods, such as GANs, variational autoencoder methods, etc. Other simplified methods, such as applying light intensity, noise, oversampling, cut-out, random erasing, etc., have improved skin cancer training by using artificial intelligence methods.

The identification and categorization of skin diseases has benefited greatly from *Deep Learning* (DL) techniques [146]. In identifying skin melanoma, several deep learning architectures have been proposed, including the ResNet residual network [142], attention-residual-learning [38], AlexNet [147], VGG [148], etc. The major difficulty in creating clinical-level AI systems is the requirement for a big dataset when applying advanced AI approaches [45, 149]. Secondly, the available datasets suffer from imbalance class, high-dimension features, noisy data, irregular sizes, and poor quality. The development of big datasets has been the focus of research; however, there are obstacles to overcome, including the lack of imaging standards and dermatology metadata, patients' confidentiality concerns, image license, detection accuracy, etc. In the event of limited data, data augmentation can aid in enhancing the accuracy of learning algorithms [150]. The most popular DAMs in skin melanoma are random picture manipulation, such as horizontal and vertical flips, random brightness, contrast, cropping, zooming, rotation, and histogram scaling, among others [151, 152]. The applications of detailed data analysis techniques must be made with a specific aim to extract the relevant features from these limited and unbalanced data.

The issue of unbalanced and tiny datasets has been addressed by using techniques such as feature space sampling, which also includes SMOTE [153], random under-sampling, clustering under-sampling, random over-sampling, etc. Fixed and random rotation angle techniques were used by [147], whereas color space transformation from RGB to HSV was researched in [123]. Thus, the use of DAM has positively impacted the performance of classifiers, improved the generalization and secure fundamental distribution of the training data. Thus, DAM can be defined as the process of enhancing the training data and the variance of particular data [88, 154]. The authors of [155] presented an improved DAM based on the covariant SMOTE technique to produce augmented data for training, enhanced the effectiveness of the classifiers, and mitigated model overfitting. Rotation and translation DAM achieved the best classification results of 99.29% [156], the conventional DAM based on zoom, shift, flipping, random transformation and rotation improved the rate to 81% in comparison with the 78% baseline result [157]; low-cost augment (LCA) strategy reached the best accuracy of 85.3% [158], the random rotation approach obtained the highest accuracy of 95.09% [159], etc. Tbl. 2.3 presents a summarized highlight of the augmentation, and classification methods, while highlighting the contributions, and shortcomings of some selected related studies as well as the application domains in computer vision tasks.

Table 2. 3. Comparison of selected related works in computer vision task

Ref.	Domain	DAM	Method	Strength(s)	Limitations
[104]	Facial Grading System	Random horizontal flip, zoom, and rotation	CNN	Overcome overfitting	Inter-rater reliability of the CNN
[160]	Facial palsy diagnosis	Geometric transformation	SVM, KNN, MLP, multinomial logistic regression (MNLR)	SVM was able to achieve better performance results in cases of partial occlusion of the face	The authors did not evaluate results with SOA methods
[59]	Facial palsy grading system	Oversampling	Machine learning algorithms	Improved performance results that could aid in joint therapeutic efforts	No further evaluation with the existing methods on the dataset
[105]	Facial paralysis classification	SMOTE	Ensemble learning algorithms	Combat issues of class imbalance	Limited dataset and lack of severity categories
[161]	Facial Paralysis recognition	Noise addition, color temperature, and rotation	MTCNN	Overcome overfitting problems and poor generalization	Limited dataset and poor labeling of the internet dataset
[162]	Person re-identification	Background substitution, geometric and color transformation	CNN	Geometric transformations significantly increase the performance	Computationally intensive
[163]	Automatic facial nerve paralysis	Translating and rotating	Combination of Parallel-hierarchy CNN and LSTM	Better detection with less processing time	Lack of various facial expressions
[102]	Facial palsy grading	Random flipping, rotation, and color jitter	3DplasyNet model	A small improvement in the performances	Computational overhead of the framework
[164]	Automatic facial paralysis detection	Generative adversarial Network (GAN)	Cascaded CNN	The study was able to successfully address the issue with limited training data	Limited public data for facial palsy studies; computationally intensive
[165]	Leaf disease detection	Flipping, zooming, cropping	ViT model based on self-attention and LeIAP	Effective pruning of unimportant attention heads and reduction of the model size,	The issue with data complexity is a result of image quality

Ref.	Domain	DAM	Method	Strength(s)	Limitations
				speed, and training time	
[166]	Plant disease identification	Axial self-attention	Inception convolutional vision transformer (ICVT)	The proposed model can acquire high-level knowledge for identifying plant diseases	Feature misrepresentation leads to misclassifications, especially when classifying diseases with similar epigenetic features
[167]	Multi-class disease detection	Random flip, cropping, random contrast, random saturation, and brightness	DLMC-Net	Reduction of trainable parameters resulting in minimal computation complexity of the model	Due to small and class imbalance, the performance is compromised for the citrus and cucumbers classes. Over-fitting of the network
[168]	Grapevine disease detection	Rotation, flipping and image brightness	Deep CNN based on EfficientNet	Reduced parameters led to an improved detection rate. The issue with the time-consuming process was overcome	Time-consuming of the proposed model
[169]	Apple disease detection	Intensity disturbance, noise addition, transformation, rotation, and flips	Deep and faster CNNs (VGG-INCEP Network Model)	Enhanced accuracy and accelerated detection speed	Misclassification of Alternaria leaf spot images due to similarity in geometric characteristics between the two disease classes
[78]	Apple disease detection	Zooming Rotation, Flipping and Shearing	ResNet, VGG (16 and 19), and Inception-v3	The proposed VGG-16 model is marginally better than the previous results	Poor generalization in fine-grained classification.
[170]	Tea disease	Deep convolutional GAN	CNN (VGG16 networks)	Significantly improved results in comparison with SVM	Accuracy is still low and requires further robustness for

34

Ref.	Domain	DAM	Method	Strength(s)	Limitations
					tea leaf disease identification
[171]	Tomato disease	Residual-based-learning	CNN	Improved performance outcome	Longer training time of 10 hours to train the model
[130]	Plant species	Translations, horizontal reflections, random addition of noise, brightness, or saturation	CNN model (ResNet and ShuffleNet-V2) models	Better classification performance; lower computational cost	Proposed method still suffers from overfitting as a result of class imbalance
[172]	Olive plant disease	Geometrical random transformation	AlexNet Architecture	The proposed method gave improved accuracy and achieved class-balancing	Proposed model training consumes a lot of time
[128]	Citrus pests and disease	Shear Rotation, Zoom, Height and Width shift	Weak DenseNet-16	With fewer parameters, Weakly DenseNet-16 achieved the maximum accuracy	Wastage of computational resources and little improvement in accuracy in comparison to SOA methods.
[173]	Melanoma detection	Enhanced Super-Resolution GAN (ESRGAN)	CNN and modified ResNet50	Enhanced automated skin cancer diagnosis technique	High computation time for training each network
[174]	Skin cancer classification	Rotating and flipping	CNN: InceptionV3, AlexNet, and RegNetY-320 models	RegNetY-320 was significantly increased	Poor generalization of the model
[175]	Skin melanoma cancer detection	Random rotation	Deep Convolutional Encoder-Decoder Net	Less overhead	The problem of uncorrelated features map
[176]	Skin lesion classification	Rotation angle and novel Gabor wavelet-	Deep CNN (AlexNet and ResNet)	Ability to isolate directional information and improve classification	Extremely low sensitivity rate when classifying melanoma
[177]	Melanoma detection	Balancing method and DCGANs	CNN (ResNet50)	The proposed method lowers computational complexity and allows resource sharing	Relatively high misclassification rate

Ref.	Domain	DAM	Method	Strength(s)	Limitations
[178]	Melanoma recognition	Rotation and pixel translation	Combined aggregated Deep Convolutional and SVM classifier	Improvement in the classification performance of the proposed method	High memory consumption and computational complexity
[179]	Skin lesion classification	Cropping, horizontal flipping and rotation angle	CNN (GoogleNet, AlexNEt, ResNet, VGGNet)	The accuracy of the system was improved. Overfitting was resolved	The method is computationally intensive
[180]	Skin lesion analysis	Cropping and horizontal flipping	CNN probability-based stepwise integration (PSI) approach (ResNet)	Visual consistency and accurate segmentation results were achieved	Highly computationally intensive
[146]	Melanoma and Naevus detection	Rotation angle and shifting	Combined Gray Level Co-occurrence Matrix (GLCM), SVM and RBF (DenseNet)	Flexibility to the variation in image quality for different images, thus reducing the misclassification rate	High computation complexity

2.5. Related Studies – Sound/Audio Application

Originally, the analysis of deep breathing sound dataset combining deep learning with image enhancement and sound spectrum technique was presented in [181].

2.5.1. Related Work in Sound Classification Methods

It is fascinating to note that the popularity of AI techniques has expanded to areas like music recommendation, speech detection, etc. [182]. Given their significance to our daily lives, and the growing demand for efficient and automated sound classification systems, the currently existing technologies for automatic sound classification are frequently used in monitoring systems [183], voice assistants [184], chatbots [185], smart safety devices [186], and in many real-world settings like engineering [187], industrial [188], domestic [189], urban [190], road [191], and natural [192] applications.

Systems for sound recognition have been developed by using ML techniques including *Random Forests* (RF), decision trees, logistic regression, multilayer perceptron, and others [193]. The performance of classification models has improved in the last decade thanks to advanced classifiers, such as DL techniques, which have shown excellent abilities to learn complicated level characteristics from original data through the removal of high-level features [194]. Recent paradigm shifts in the advancement of deep neural network performance apply dropout, regularization, fine-tuning hyperparameters, momentum gradient descent approaches, etc. [195–198].

CNNs, which are a widely used deep network model, were able to produce considerable and more accurate training outcomes, and sound classification approaches have demonstrated an easy shift from basic ML classifiers to sophisticated deep learning classifiers [199]. However, this method struggles with poor performance since there is not enough data available to address audio/sound-related problems, loud-audio-signals, and industrial sounds [200, 201]. Numerous researchers have been captivated by the idea of using various ML algorithms in sound categorization because of the extensive applications of deep learning techniques and diverse architectures [202–206]. Other proposed DL architectures include MobileNet, NasNetmobile, ResNet18, ResNet50, and ResNet101 [207]. CNN, LSTM, and Resnet50 [208], ResNet50 and MobileNet network [209]. CovScanNet [210], the lightweight deep learning model [211], the Shallow ML algorithm, and CNN [212, 213].

Although deep learning networks have produced some amazing results in the categorization of sounds and audio, there are still some issues with the presently existing methods, such as poor results scenarios caused by additional factors such as reverberations, noise types, channels, etc. [214], inaccurate annotation [215], or deficiencies in audio feature representation [216]. The introduction of DAM methods has significantly helped in combating some of these limitations with notable outcomes. Examples of DAM approaches applied to sound data include a combination of stretching, shifting, dynamic range compression and background noise increasing accuracy to +5% [217], the integrated image and audio traditional augmentation approach was able to obtain the best accuracy of 99.04% when using ResNet152 [206]. Filter-Augment enhanced detection with an increment of +6.5% [218];

WaveGAN improved the detection rate with increased the accuracy of +2.31% [219], shuffling and mix-up augmentation obtained an overall F1-score of 89.95% in comparison with the baseline result of 85.50% [220]. The summary of the categories of the literature based on the dataset, feature extraction approaches, classification methods and data augmentation methods used in sound classification tasks is presented in Fig. 2.4. In addition, the summary of selected studies on sound classification and their comparison based on methods, contributions and limitations are illustrated in Tbl. 2.4.

Sound Classification: Sources/ Methods/Metrics

Sound Databases

ENVIRONMENTAL SOUND DATABASES
- UrbanSound8k
- ESC-10
- ESC-50
- DCASE-2017
- Archive RecognitionDB
- TUT Urban Acoustic Scenes 2018 dataset
- NIST SRE'16
- Acoustic Scenes 2019 dataset
- SINS dataset
- Mivia Audio Event Dataset
- Mivia Road Events
- YBSS-200

SPEECH DATABASES (Language)
- THCHS30 dataset (Chinese)
- AISHELL dataset (Mandarin)
- DIRHA project (English)
- HSC14
- EVALITA (Italian)
- WSJCAM0 adults' speech corpus (UK-English)
- PF-STAR children's speech corpus (English)
- Indonesian Ethnic Speaker (Indonesian)
- Aurora-4 (Finnish)
- British English speech corpus
- OY14-CEN0 Mandarin-English mix lingual speech data
- VCTK corpus (English)

EMOTIONAL SOUND DATABASES
- RAVDESS
- Toronto Emotional Speech Set (TESS) VOCE
- SAVEE
- Acted Emotional Speech Dynamic Database
- eNTERFACE
- Berlin Database of Emotional Speech
- EmotionDB 1999
- Emotional Soundscapes database

MEDICAL-SOUND DATABASES
- ICBHI-17 dataset
- 2016 PhysioNet CinC
- PASCAL Heart Sound Challenge dataset
- COVID-19 sound dataset Cambridge University
- Dysarthric Speech Corpus in Tamil
- Store Sound Corpus (LPSSC)
- Universal Access research (UA-corpus)
- Gastrointestinal Sound Dataset

MUSICAL-SOUND DATABASES
- GTZAN
- Ballroom
- Homburg
- ISMIR2004
- INTERSPEECH 2017

ANIMAL SOUND DATABASES
- BIRDZ
- CAT
- DCLDE dataset
- NABW calls dataset

EVERYDAY-SOUND DATABASES
- Freesound
- MSOS
- AudioSet
- Sound Events for Surveillance Applications (SESA)

SYNTHETIC-SOUND DATABASES
- RADIOAGL 2016 04C dataset
- ACARS dataset
- Real Word Computing Partnership (RWCP)
- CMU arctic database

Feature Extraction
- Mel-Frequency Cepstral Coefficients
- waveform
- Melspectrogram
- Cepstral mean and variance normalization (CMVN)
- multi-resolution feature extraction
- RGB Spectrogram
- Spectral Bandwidth
- Log-Mel spectrogram
- Inverted MFCC (IMFCC)
- STFT
- Spectral Centroid
- Spectral Contrast
- spectrogram
- STFT and IFFT
- Spectral Flatness and Rolloff
- zero crossing rate (ZCR)
- chroma deviation feature
- spectral centroid spectral slope
- entropy of energy
- RedAV
- Constant Q transform (CQT)
- Fourier transform magnitude (STFTM) spectrum
- Gammatonegram (GCC)
- Mel-Frequency Cepstral Coefficient and Gammatone Frequency cepstral coefficient (MF-GFCC)
- recursive feature elimination (RFE)
- log-gammatone spectrogram
- spectral roll-off
- GCC-PHAT Pattern features

Audio-based Augmentation
- Pitch Scaling
- Time Shifting
- Noise Injection
- Dynamic Range Compression
- Random Masking
- Amplitude Interference
- Overlapping windows
- MixUp
- Vocal Tract Length Perturbation (VTLP)
- Equalized Mixture data Augmentation (EMDA)
- Speed Perturbation
- oversampling (SMOTE)
- Latent Dirichlet Allocation (LDA)
- Channel Distortion
- Trimming

Data Augmentation Methods

Standard Image Augmentation
- Multi-resolution
- Linear Interpolation
- SMOTE
- Flipping
- Random cropping
- Greyscale
- Zoom
- Width Shaft
- Noise
- Label smoothing mixup
- Random Image warp
- Rotation
- Random Erasing

Advanced-Image Augmentation
- Generative Adversarial Network (GAN)
- Auxiliary classifier generative adversarial network (AC-GAN)
- Weighted Cycle-Consistent Generative Adversarial Network (WCGAN)
- Generative adversarial network (GAN) and variational autoencoder (VAE)
- Semi-supervised Conditional Generative Adversarial Networks (scGAN)
- WaveGAN
- Zoom Impulse Response generator

Evaluation Metrics
- Unweighted Average Recall (UAR)
- Equal Error Rate (EER)
- Accuracy
- Recall Sensitivity
- Precision
- Specificity
- F1-Score
- AUC
- Recognition Rate (RR)
- Token Error rate (TER)
- Mean squared error (MSEs)

Fig. 2. 4. Categories of methods, features, and augmentation techniques used in sound classification

39

Table 2. 4. Comparison of selected related works in the Sound/Speech classification task

References	Feature Extraction	Augmentation techniques	Classification Method	Strength(s)	Limitation(s)
[2]	MFCC and Zero crossing rate (ZCR)	Multi-domain integration (MDI), Noisy data augmentation (NDA), and Multiple SNR consolidation (MSC)	MA-CapsNet model	Computational efficiency in real-time SER and also improvement of the learning capabilities to nonlinear separability of speech data	The misclassification rate is high for some emotions due to the class proximity of emotions in arousal and valence dimensions
[201]	MFCC, log-mel-based model, waveform based	Mix-up approach (linear interpolation) and scale augmentation	CNN (DenseNet module)	The study proposed a fast and high-performance system for effective recognition of musical instruments	Real-world instrument sounds are not constant and thus differ in temporal, spectral, and dynamic features and could result in a misclassification rate
[208]	Mel spectrogram (Mel), Mel Frequency Cepstral Coefficient (MFCC) and Log-Mel	Pitch shifting, silence trimming, time stretch, adding white noise	DCNN	Best results for the environment sound classification by using augmentation on Log-Mel	Computational time complexity
[210]	Mel spectrogram	Generative adversarial network (GAN) and conventional method.	Deep convolutional neural network	Significantly better performance	The outcome of the basic deformation (pitch shift and time shift) degrades the performance
[223]	Short Time Fourier Transform (STFT)	Random noise, Spec Augmentation	CNN (GoogLeNet and AlexNet)	Significant improvement in the detection rate for minority class (unhealthy heart sounds)	Generalization issues and issues with noisy datasets resulting in poor performance of the model

References	Feature Extraction	Augmentation techniques	Classification Method	Strength(s)	Limitation(s)
[224]	zero crossing rate, energy, entropy of energy, MFCCs	Random oversampling	Random Forest and DCNN	Increased prediction performance on limited datasets	Insufficient data
[225]	MFCC	Multi-resolution feature extraction (MRFE)	Hybrid deep neural network (DNN)-hidden Markov model (HMM)	Proposed data augmentation presented the least WER when compared to existing methods	Difficulty in training low and very low intelligible category of dysarthric speakers
[226]	STFT and IFFT	Random cropping, label, interference (spectrum and amplitude)	Deep CNN	Improved classification stability and prediction performance of the modulation scheme	High memory computational and lack of sensitivity to ambient noise
[227]	Constant Q transform (CQT)	Auxiliary-GAN, shifting and stretching	CNN	Reduced overfitting and improved generalization	Time and computational constraints and misclassification errors due to sound similarity
[228]	MFCC, inverted MFCC, Mel, and Bank filter bank-derived features	SMOTE	SVM, RF, REPTree and MLP	Improved detection performance for some of the datasets. The study showed consistent and optimized performance in all cases	Misclassification of models based on categorical emotion class, especially in classes of fear and happiness
[229]	MFCC	Reverberation and noises (MUSAN, Babble and static)	DNN autoencoder	Robust and universal technique based on its stability	Poor performance of the model due to due to language mismatch

2.6. Gaps in Literature and Research Opportunities

The contributions of artificial intelligence methods in small data analytics have continued to show promising results, as shown by the literature. These success rates can be attributed to the adoption of several mechanisms, such as the application of regularization methods like label smoothing, dropblock, etc. However, some of the existing approaches deployed for improving the performance of learning classifiers still suffer from significant challenges such as:

1. The huge dependency on big datasets in AI applications affects the performance of machine learning-based models. Therefore, there is a need to reduce the overall dependence on big data for training robust models.

2. The complexity of sufficiently identifying relevant features due to variations in the properties of disease patterns, especially in image classification tasks for disease recognition has been manifested. This is a major problem associated with some of the existing augmentation techniques thereby affecting generalization and causing poor recognition results of the classifier.

3. The application of an unconstrained conventional method, especially the use of elastic transformation, produces over-augmented or unrealistic images; however, the significance of the impact of these is still debated in the literature. In addition, the adoption of erasing approaches suffers from a lack of smoothness, multiple structural breaks, and inconsistent features within augmented images. Thus, generating synthetic image samples that differ completely from the training data or the target data has no significant influence on the performance of learning models.

4. Some of the advanced augmentation methods, such as GANs, generate unrealistic and inapplicable synthetic images and do not necessarily improve the generalization of the learning model. Thus, it is important to use augmentation strategy methods which would enhance the feature learning capability of the model.

5. The high computational complexity of the majority of the existing deep learning models is based on the use of the same feature weights applied across different levels. However, to develop a more robust classification model, it is important to adopt an advanced augmentation model to transfer learning for better knowledge to deeper layers of the network for optimal classification.

To address the research gaps highlighted above, this dissertation shall employ various data augmentation methods in the next chapter with the major consideration in the peculiarity of the feature representation and the classification problem to be addressed.

2.7. Review Summary

The classification task irrespective of the research domain suffers immensely from the problem of access to reliable data. Common knowledge from the literature has shown that, in order to obtain high generalization capabilities, CNN-based models

require more extensive training data. If the deep learning models can extract valuable data from the original dataset, these models can perform better even while using poor-quality data. However, producing massive datasets is expensive, time-consuming, or almost impossible to collate. Data augmentation plays a crucial role in producing promising ways to improve the accuracy of classification models in small data analytics (images, sound, time series, etc.) and helps to overcome the overfitting of learning models. By injecting more training data samples into the training datasets, this technique has been used to solve certain challenges, particularly in the field of computer vision. Additionally, data augmentation is essential to amass sufficient data so that to achieve adequate performance and handle the problems associated with data scarcity. Examples of data augmentation approaches which have been used in increasing and balancing datasets are random cropping, principal component resampling, scaling, adding noise, Gaussian blurring, rotation, flipping images, color jittering, etc. It has been shown that this approach is particularly successful in learning from small and unbalanced class datasets. Applying augmentation techniques to a small number of samples inside a very high-class imbalance, however, might not produce the desired variances for generating different samples for rebalancing the dataset.

Access to reliable datasets, overfitting of learning models, poor generalization of models, and the need to build efficient classifier and detection algorithms by using small data are only a few of the issues which make data analytics complicated. Background noise on sound datasets is a problem for data augmentation in most of the image and sound classification tasks. Effective feature extraction is impacted, and the classifier performance would be affected if noisy samples were turned into synthetic datasets. Therefore, the obvious solution is to propose a successful data augmentation model that can produce high-quality simulated data with little loss, while also enhancing the effectiveness of learning models in the process.

In Chapter 3, the dataset applied to image classification and sound classification tasks shall be discussed in detail. The theoretical data augmentation techniques and the algorithms shall be addressed in Chapter 4, while Chapter 5 presents the evaluation results achieved in the application of data augmentation techniques in training and validation of small datasets in images, and sound classification scenarios.

3. DATASETS USED FOR THE RESEARCH

In this chapter, the datasets used in the studies were introduced. Data types were distributed into different subsections, namely, image database and sound databases, as shown in the subsections.

3.1. Image Classification Databases

3.1.1. Description of YouTube Facial Palsy (YFP) Dataset

The facial palsy dataset using the publicly accessible *YouTube Facial Palsy* (YFP) video was collected by [99]. This YFP dataset contains only Bell palsy videos, and it was collated for the identification of facial palsy symptoms. It consists of 32 videos from 21 patients. The labelling/annotation was done by three medical specialists. The patient's facial expression shows the palsy deformation with a variant in severity image patterns captured by the camera across variations of time. All the videos were converted into an image sequence, and a total of 1105 palsy facial images were extracted with a sampling rate of 6 fps. The facial images are front portrait views with a resolution of 227×227 pixels, and each image is a distinct shot using various facial expressions, illumination intensity, and background.

The non-palsy (or healthy) images were obtained from the *Caltech Face Database* [223] which consists of 450 facial images obtained from 27 distinct individuals under various lighting conditions, with varying facial expressions and surroundings. The raw image size has a resolution of 896×592 pixels, which was pre-processed by resizing to an acceptable deep model resolution of 227×227 pixels. Therefore, the total images collected comprise 1555 facial images which are categorized into 1105 palsy-positive images and 450 non-palsy images. Fig. 3.1 depicts instances of face palsy and normal image classes, whereas Table 3.1 shows the prior methods of studies based on YFP images.

Table 3. 1. YFP dataset classification methods in related work

References	Methods	Objective/Area
[99]	Deep Hierarchical Network (DHN)	Intensity variation
[163]	Hierarchical network with LSTM	Severity grade classification
[160]	Machine classifiers such as KNN, SVM, MLP, and multinomial logistic regression (MNLR)	Face palsy detection and Intensity variation
[224]	CNN and nine ML algorithms	Bell palsy and Eye-blinking detection
[225]	Semi-supervised extreme learning machine (RC-SSELM)	Facial nerve paralysis (FNP)
[226]	Principal component analysis with support vector machine	Facial paralysis recognition

Fig. 3. 1. Instances of palsy and normal face images. (Row A) palsy face images (Row B) normal face images [89]

3.1.2. Cassava Disease Dataset

The Cassava dataset consists of cassava leaf images. It was gathered by the AI laboratory of the Makerere University and the National Crops Resources Research Institute, Kampala, Uganda [108]. This dataset was majorly crowdsourced from farmers, and it comprises four fine-grained cassava leaf disease categories which are *Cassava-Mosaic-Disease* (CMD), *Cassava-Bacterial-Blight* (CBB), *Cassava-Green-Mottle* (CGM), *Cassava-Brown-Streak-Disease* (CBSD), and healthy cassava leaf images. A total of 9436 data samples is available in the database, and it comprises annotated data of 5656 samples and 3774 unlabeled images where the former contains 2658-CMD, 1443-CBSD, 773-CGM, 466-CBB, and 316-healthy images, respectively. This dataset is made up of five classes, namely, healthy, CBB, CMD, CBSD, and CGM, respectively. A sample of each of the five kinds of cassava leaves is shown in Fig. 3.2.
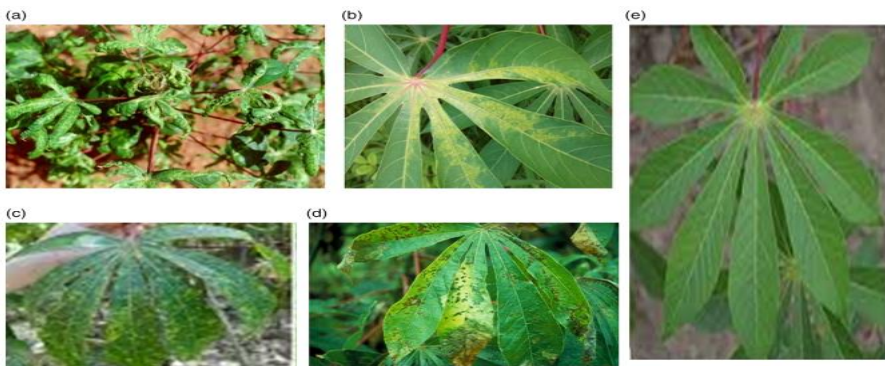
Fig. 3. 2. Image instances of cassava dataset classes: (a) CMD (b) CBSD (c) CGM (d) CBB and (e) healthy

A chlorotic ring surrounds the angular brown patches that are indicative of CBB on the leaves, whereas CMD is distinguished by a mosaic of chlorotic leaves and leaf deformation. Infected leaves with CBSD exhibit significant chlorosis and necrosis, which turns them yellow. From a few scattered spots to a full loss of chlorophyll, CGM causes yellow discoloration on leaves. The description of the cross-validation ratio and the number of sample instances is depicted in Tbl. 3.2, and Tbl. 3.2 represents summarized methods and references of the relevant literature.

Table 3. 2. Cassava disease dataset sample distribution in 5 classes

Classes	Training samples (#)	Testing samples (#)
CMD	2126	532
CBSD	1154	287
CGM	618	155
CBB	373	93
Healthy	253	63

Table 3. 3. Cassava disease dataset classification methods in related work

References	Methods	Objective/Area
[227]	Enhanced CNN models (ECNN)	Cassava Leaf Disease Identification
[228]	Bi-LSTM classifier	
[229]	Transformer-embedded ResNet (T-RNet) model	Cassava Leaf Disease classification
[230]	KNN algorithms	
[231]	ArsenicNetPlus	Cassava disease detection
[232]	Deep residual convolution neural network	CMD disease detection

3.1.3. Description of PH2 Dataset

The PH2 dataset was gathered by the Dermatology Service of Hospital Pedro Hispano, Portugal, and it comprises 200 images in total, with samples representing 80, 80, and 40 instances of each of the three clinical diagnosis classes, namely, common nevus, atypical nevus, and melanoma, respectively [141]. This dataset consists of JPEG and PNG formats for skin lesion images. The dataset is partitioned into training and test image sets, each of which includes images and ground truth labels. In this study, the 60:40 cross-validation ratio was applied with 60% for the training data and the remaining 40% for testing the trained network. After employing the proposed augmentation techniques, two different databases were developed. The covariant SMOTE technique which was used to create synthetic data is referred to as *AugDB-1* in the next chapter. The covSMOTE approach was utilized to balance the melanoma (minority-class). For the second augmentation approach, a standard geometric augmentation approach was used, such as flipping, rotation, translation,

etc., and the new database shall be referred to as *AugDB-2*. Tbl. 3.4 and Tbl. 3.5 provide a data sample highlight. The outcome of our proposed model on the different data types is categorized as binary and multiclass classifications. The analytical results shall be presented in the next chapter.

Table 3. 4. PH2 dataset sample distribution in 3 classes

Classes	Total No. samples	Training samples (#)	Testing samples (#)
Melanoma	40	24	16
Atypical Nevu	80	48	32
Common Nevu	80	48	32
Total	200	120	80

Table 3. 5. PH2 dataset classification methods in related work

References	Methods	Objective/Area
[233]	CNN models	Melanoma detection
[234]	Eight pre-trained CNN models	Skin cancer diagnosis
[235]	WT-DRNNet	Skin lesion classification
[236]	17 pre-trained CNN models	
[237]	DL architectures based on Res-Unet	Skin lesion segmentation and classification
[238]	Full resolution CNN (FrCN)	
[239]	U-Net and a Fully CNN (FCN8s)	
[222]	Deep CNN model	Classification of skin cancer

3.2. Sound Classification Databases

3.2.1. Overview of the COSWARA Dataset

The COSWARA Dataset was created by [240] to provide a cost-effective tool for the COVID-19 diagnosis by using cough, speech, and breath sounds. This dataset was targeted towards the detection and measurement of biomarkers of this disease in the acoustic content of these sounds. The COSWARA database contains 2130 recordings for nine different categories which are: breathing (deep and shallow), cough (heavy and shallow), digit counts (fast-paced and normal), and vowel phonation (/ey/, /i/, /u/). In addition, other metadata are also included in the database, such as the gender, age, location, health status, and the presence of co-morbidity. There are seven audio files in the COSWARA dataset, and it is summarized in Tbl. 3.6 for each category. In this study, the focus was majorly on the deep breathing audio samples (coined as COCOA-DB) and four major classes are experimental for COVID-19 classification. The highlights of the methods using the COSWARA dataset are presented in Tbl. 3.7.

Table 3. 6. COSWARA dataset sample distribution in 7 classes

Data Classes	Size	No of Audio samples (#)
Positive_Asymptotic	48kHz	42
Positive_Mild	48kHz	231
Positive_Moderate	48kHz	72
Recovered_full	48kHz	99
RINI (Respiratory illness not identified)	48kHz	150
NRIE (No respiratory illness exposed)	48kHz	164
Healthy	48KHz	1372

Table 3. 7. COSWARA dataset classification methods in related work

References	Methods	Objective/Area
[241]	Ensemble deep learning algorithms	COVID-19 classification
[242]	Machine learning algorithms	COVID-19 cough classification
[243]	Ensemble-based multi-criteria decision making (MCDM)	COVID-19 detection
[212]	Shallows machine learning, CNN models	
[244]	LR, Gradient Boosting trees and SVM	
[245]	SVM, LDA, kNN, and partial least squares regression (PLSR)	COVID-19 classification

3.3. Summary of Dataset used

Data is the most relevant part which influences the performance of computer vision applications, sound applications area, etc. This chapter explores the different publicly available datasets used in this dissertation for different classification tasks. The datasets have been carefully examined and described concerning the application areas. The dataset used for the computer vision tasks contains images of real-world scenes, such as the YFP dataset (facial palsy data), Cassava disease data, and PH2 dataset (skin melanoma images). For the sound classification task, the COSWARA dataset was analyzed. The datasets are a continuation of the research done by previous researchers in published academic articles.

4. METHODOLOGY

4.1. Overview of Research Methodology

In this chapter, detailed discussions of the overview of the proposed data augmentation methods and frameworks used for evaluating different classification tasks shall be addressed. To further analyze the theoretical backgrounds of the proposed augmentation techniques, three categories of augmentation applications shall be presented which are: image augmentation, sound augmentation, and numerical-based augmentation. Fig. 4.1 depicts the scheme or flow of the proposed study methodology.
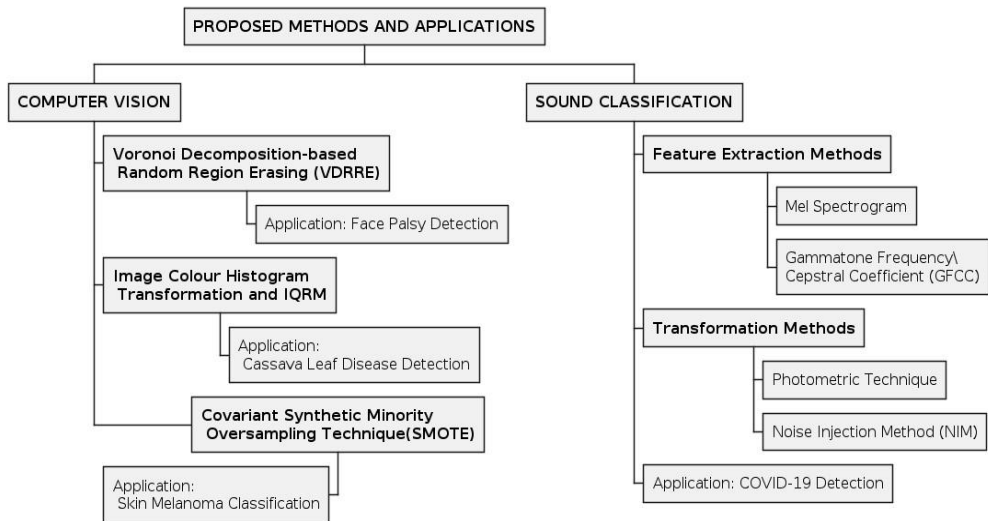
Fig. 4. 1. Scheme of the proposed methodology

4.2. Motivation

There are various methods proposed in previous studies for reducing over-generalization and overfitting of learning models, such as the use of regularization approaches, lowering the learning rates, modifying the loss functions (applying label smoothing and weight decay), dropout, etc. However, the application of data augmentation, or transformation methods, has improved model generalization in small data analytics to a great extent. The main goal of DAM approaches is to improve the performance of training models by mitigating overfitting. The application of augmentation or transformation methods is suitable in data analytics as the general problems affecting most publicly available datasets are as follows: (1) insufficient data, and (2) imbalanced classes of data.

The quality of the input-data representations alone determines how well ML algorithms perform. The idea of small dataset learning is to develop enhanced augmentation techniques with the ability to extract the relevant features/information from this small sample; therefore, the idea is to reduce the complexity and pilot run times. According to the statistical learning theory, generalization can be improved by determining a model's effective complexity, such as precisely setting the regularization; nonetheless, generalization should always get better with more training

samples [53]. Furthermore, medical datasets suffer majorly from limited data samples due to some constraints such as privacy of patient data, costs of data gathering, the vigor of expert intervention in annotating the datasets, etc. Therefore, the application of effective augmentation methods is extremely important in minimizing the generalization error and significantly improving the performance of the learning models.

4.3. Justification of the Proposed Methods

Face Palsy Detection: The *Voronoi-Decomposition-based-Random-Region-Erasing* (VDRRE) technique has been selected to augment the data in face palsy recognition to overcome the lack of variation in facial expressions within the dataset. To recognize face palsy, it is necessary to capture a diverse array of facial expressions and unique variations that are characteristic of individuals suffering from face palsy. The utilization of the VDRRE technique will allow for the creation of artificial facial features by randomly occluding some parts of facial expressions. This will generally enhance the precision and applicability of the recognition models.

Skin Melanoma Detection: The *Covariant SMOTE* (CovSMOTE) technique is used based on its inherent ability to tackle the issue of imbalanced datasets. Imbalanced datasets are frequently encountered in medical fields, where the prevalence of certain conditions is comparatively low. Thus, the CovSMOTE method is used considering its capability to develop synthetic samples of the minority class while considering the covariance structure of the data. The application of this method will improve the class balance and thereby enhance the overall performance and generalization of deep learning models in skin melanoma detection.

Cassava leaf disease detection: The novel augmentation method based on the combination of the *Image Color Histogram Equalization* and *Image Quality Reduction* techniques will be implemented for cassava disease detection. The choice of this method is based on the consideration of colors, illumination, and texture as an effective attribute in determining the plant leaf quality and its health status. This method will incorporate the diversities of visual representations of cassava leaf disease images, thus simulating distinct lighting conditions, image qualities, and levels of noise to generate synthetic image samples. The implementation of these variations into the dataset is expected to aid the deep learning models to get trained better on the augmented data with the possibility of acquiring better recognition of different cassava leaf disease types in various environmental circumstances.

Although each proposed augmentation approach is performed differently in different broad experiments, it is still possible that a single technique could be suitable in specific situations. However, it would not fully capture the distinct characteristics and nuances found in each specific area. Thus, employing these various techniques enables the customization of the data augmentation procedure to suit the specific requirements of each research field, which would ultimately result in more accurate and resilient deep learning models in each specific domain. In the next section, the detailed illustration of each proposed augmentation method is fully discussed.

50

4.4. Proposed Image Augmentation Methods

The proposed image augmentation methods used in these studies are divided into five distinct categories, as discussed in detail in the subsection below.

4.4.1. Proposed Voronoi-Decomposition-based-Random-Region-Erasing (VDRRE)[1] – Face palsy detection

The overall methodology structure proposed for the face palsy detection is depicted in Fig. 4.2. It is categorized into five parts. The section discusses the method phases beginning with a few-shot learning approach by using one-shot and two-shot learning approaches. A detailed explanation of all the steps and procedures involved is discussed in this subsection and presented in Fig. 4.3.
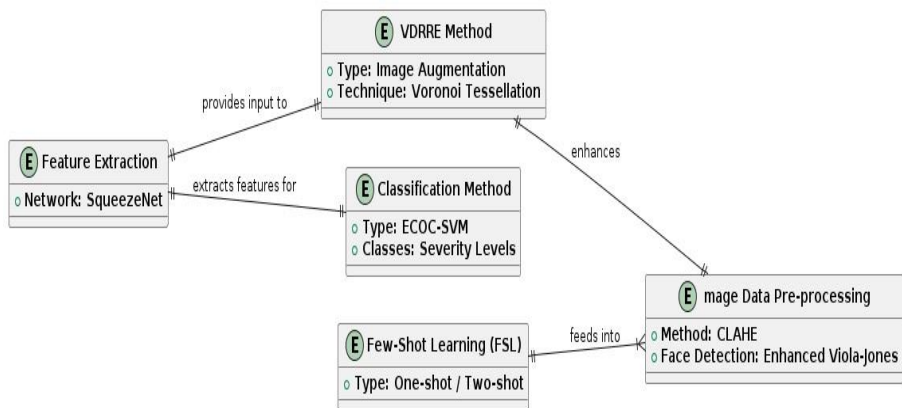
Fig. 4. 2. Entity diagram of the proposed VDRRE method

There are five stages involved in this study which are listed below, and the flow diagram of the face palsy detection showing the proposed VDRRE, Geometric and Random-erasing augmentation is depicted in Fig. 4.3.

---

[1] The Material/Content presented in this section has been published in [89]
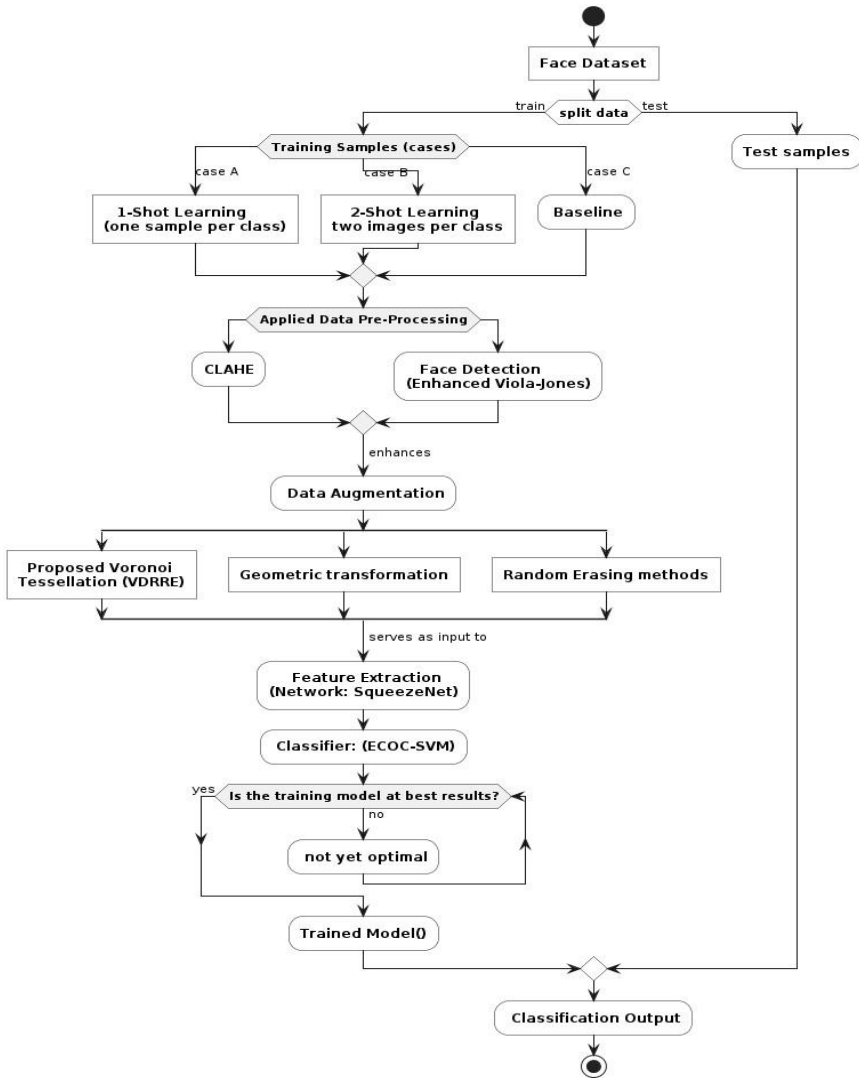
Fig. 4. 3. Activity UML-diagram of the face palsy detection showing the proposed VDRRE, Geometric and Random-erasing augmentation

1. **Few-Shot Learning (FSL):** The training sample is based on the few-shot learning method. This is a learning method which tries to emulate human intelligence based on small data sizes or single training items. The FSL method is inspired by the special intelligence of humans to accurately generalize after observing one instance of a given object. Contrary to the normal practice of using big data for training, FSL only attempts to use very small data in a learning model. Given a small number of input-output sample pairs, FSL trains a classifier h, where output y_i represents the class (label) of the independent variable x_i. In this study, two instances of FSL were used,

52

and it is based on the same principle which served for *One-Shot-Learning* (OSL) and *Two-Shot-Learning* (TSL). OSL is a severe example of few-shot learning [98], wherein the classifier can be trained by using merely one instance with a class label. The objective of few-shot classification is to minimize the detection error on unlabeled data. Suppose dataset $d \in D$ comprises pairs of features and labels $\{(x_i, y_i)\}$, whereby each label is part of a recognized set of labels L. The training S and testing B samples constitute dataset d, which is partitioned into two parts $d = (S, B)$. The training set has K -labeled instances for each of the N classes, and accepts a K -shot problem with N classes by using the steps outlined below:

    a.   Select a subset of the labels., $T \subset L$.

    b.   Select a training set $S^T \subset D$ and a testing set $B^T \subset D$ (consist of data with labels from the subset from item) 1: L, $y \in L$, $\forall (x, y) \in S^T$, B^T.

    c.   The set of S^T is presented as input of the model.

      To compute the model parameters and the loss function, backpropagation in the final optimization was used in addition to numerous testing sets B^T, much like in supervised learning.

2. **Image Data Pre-processing:** *Contrast Limited Adaptive Histogram Equalization* (CLAHE) was applied to transform all colored facial images to grayscale to remove the undesired color contrast. By using CLAHE, the histogram was clipped at a predetermined value, and contrast amplification was reduced, hence improving noise removal, shadows, and light instability. In this study, an enhanced CLAHE approach was applied to substitute the neighborhood conditional histogram for the clip-redistributed histogram [246]. By using this approach, the local contrast was optimized and, therefore, the facial images were improved based on the edge information. To diminish the block artefacts and improve the local contrast, both global and local mapping functions were combined with the neighborhood conditional histogram, and a typical sample of the contrast enhancement result is presented in Fig. 4.4.

Fig. 4. 4. Sample of contrast enhancement: (A) original image; (B) contrast-enhanced image

An enhanced Viola-Jones face detection approach was implemented by using the Haar-like rectangular feature expansion [247]. Image resampling and 2D convolution separation methods were combined in this detection system. Shape and edge, template matching, and face feature models are combined by the traditional Viola-Jones approach using classifier boosting. Prior to doing further feature assessment on the integral image, the Haar-like feature matrix is used to scale face features. The AdaBoost algorithm [248] was utilized to create cascade classifiers and stronger classifiers to eliminate non-face images and boost the accuracy. A rotatable 45° rectangle feature was applied to compute the values of the integral images rather than using the conventional Haar-like orthogonal features. Image traversal was used to achieve the total pixel of the image regions and to eliminate unwanted features; hence, a cascade of weak classifiers was employed to successfully identify a face from the images. These cascaded weak classifiers were able to create one classifier while the window was sliding across the whole image. An instance of face detection is depicted in Fig. 4.5.

Fig. 4. 5. Face detection features: Rotatable 45-degree (left); Input image (middle); and Face detection outcome (right)

3. **Proposed VDRRE Method:** For all classes, a single-shot learning sample was enhanced with random partition erasing image augmentation. The VDRRE approach was introduced, and it is a novel approach for image enhancement. 2D plane images were divided into areas near each of a specified set of points and the coordinates of these points were produced by using arbitrary numbers that are taken from uniform distribution. Basically, we establish a partition in the image by selecting N randomly distributed points that are Voronoi tessellation and thus split an image into the area of Voronoi around each specific set of objects [249]. Take, for instance, a set of generators $P = \{p_1, p_2, \dots p_n\} \in R^2$, where the distance from any area of X in the plane to the generator point P_n is represented by $dist(X, p_n)$. Choose the closest generator $p_n \, \epsilon P$ with a specific distance metric $dist$ for all potential locations of X in S. Subsequently, if it is within proximity of two generators in P, the interval turns into an edge; alternatively, if it is near more than two generators, the position changes into a vertex. Let distance (p, X_i) represent the Euclidean distance between any given position p in the space and the

primitive area X_i. Equation (4.1) can be used to determine the bisector between X_i and X_j, and Equation (4.2) can be used to indicate the region where X_i dominates X_j:

$$b(X_i, X_j) = \{\text{dist}(p, X_i) = \text{dist}(p, X_j)\}, \qquad (4.1)$$
$$\text{Dom}(X_i, X_j) = \{\text{dist}(p, X_i) \leq \text{dist}(p, X_j)\}, \qquad (4.2)$$

$b(X_i, X_j)$ is the perpendicular bisector of the line connecting X_i, X_j, $\text{Dom}(X_i, X_j)$ is the dominance region of X_i over X_j. Here $i \neq j$ and $\text{dist}(p, X_i)$ is the weighted distance between p and X_i. The Voronoi region for a simple X_i can be defined as follows in Equation (4.3):

$$V(X_i) = \cap_{X_j \in X \backslash \{X_i\}} \text{Dom}(X_i, X_j), \qquad (4.3)$$

$V(X_i)$, is called the weighted Voronoi region. As a result of this assignment, an image is decomposed into several Voronoi cells with image boundaries. This process is known as Voronoi tessellation, and the flow diagram of the activities is depicted in Fig. 4.5. Tbl. 4.1 shows the algorithm for VDRRE. Lastly, to complete the occlusion and create a new image, an irregularly selected Voronoi cell is full of randomly selected pixel color values obtained from an equivalent distribution, as demonstrated in Fig. 4.7.

Table 4. 1. Algorithm for VDRRE

Algorithm 1: Voronoi decomposition random region erasing
Inputs: image: input image M,
       Image size: X and Y
Output: Voronoi image M*
Processes Augment (image, P, v)
       Area of polygon that surrounds all points in the image: S
       Number of points: N
       Generators: set of seeds for Voronoi decomposition G
       Voronoi random region: V_i
       Voronoi-erasing aspect ratio range c_1 and c_2|
       $x_{img}, y_{img} \leftarrow$ get ImageSize (M)   - Image width and height
       $x_{in}, y_{in} \leftarrow$ getAugInputSize   - model input width and height
       local generator set, G (p, q, r)   - local generator sets: points inside Voronoi cell
       For $x_{img}, y_{img} \leftarrow$ get ImageSize (M) do;
             $M == [x_{in}, y_{in}]$ do;
             Voronoi cell: $\{V_i\}_{i=1}^G$
             Compute random centroids: centroids $\leftarrow$ Rand $((c_1, c_2), [N\ 2], \text{seed}) * x_{in}$);
             Compute Voronoi diagram: vmask $\leftarrow$ v(centroids(:,1), centroids(:,2),$[x_{in}, y_{in}]$);
             Generate a random Voronoi cell within the image size X and Y.
             $P_e \leftarrow$ Rand $(P_a, P_b) \times$ P;
             Add noise to mask M* $\leftarrow$ gauss.Filter(vmask);   - noise is added to the erased area
       end
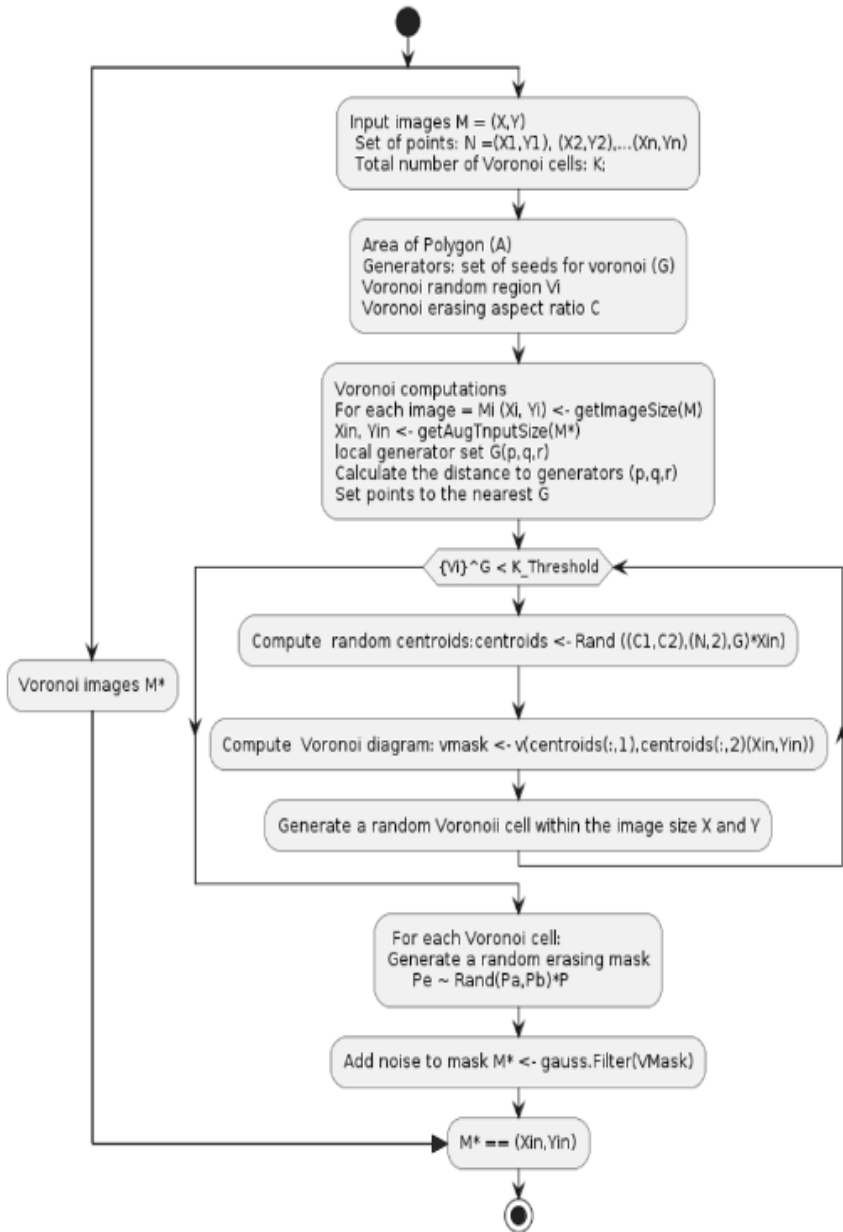
Fig. 4. 6. Flow diagram of VDRRE

From Fig. 4.7, it can be deduced that Voronoi tessellation produces polygons with more intricate shapes than the commonly utilized plain rectangular ones. Our novel image augmentation method was able to generate more synthetic images thus developing a more reliable classification model by using images

to increase the training dataset depending on various degrees of occlusion while eliminating the risk of overfitting.
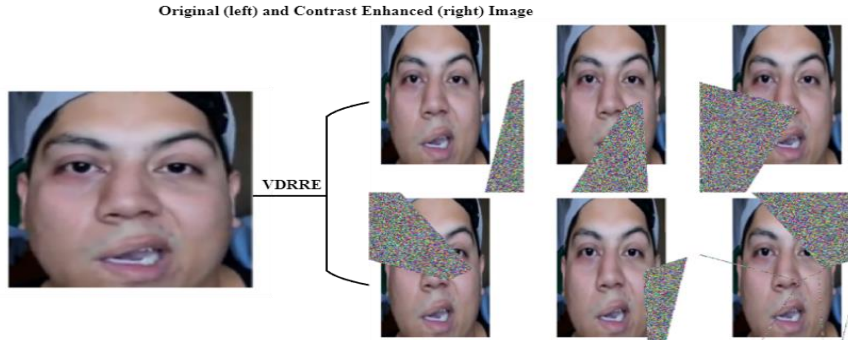
Fig. 4. 7. Original image (left); Augmented image generated from the proposed random region erasing method (right)

4. **Classification Methods and Feature Extraction:** *SqueezeNet* pre-trained network [250] was applied to extract pertinent features by using the ImageNet dataset. The choice of SqueezeNet against other architectures (such as Inception, AlexNet, VGG, ResNet, etc.) is based on the fact that SqueezeNet is capable of working with limited files lower than 0.5 MB and with insufficient input image sizes with no need for resizing. Although, the standard image size of 224×224 is still the best case. The major operating framework of the SqueezeNet network includes the fire unit consisting of *Squeeze-Layer* (SL), *Expand-Layer* (EL), and some *Pooling-Layers* (PL). The squeeze layer minimizes the feature map sizes; the expand layer is applied to expand the layers; it is depicted in Fig. 4.8.
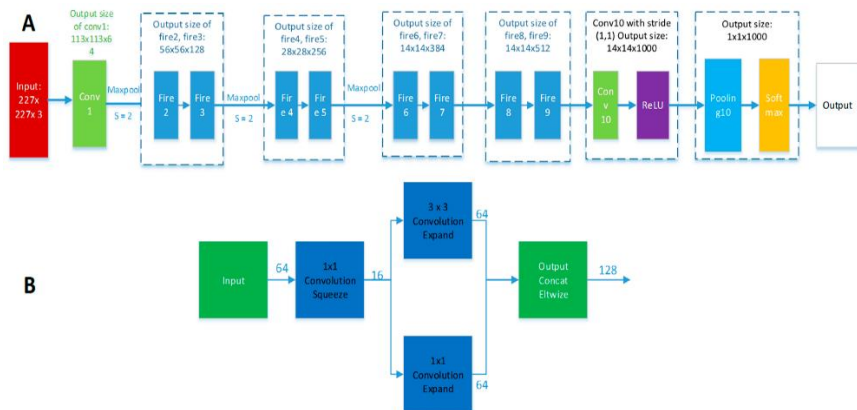
Fig. 4. 8. SqueezeNet architecture. (A) Complete view; and (B) Fire unit with squeeze and expand layers

To achieve an increased abstraction level, there is a need to increase the total filters by using a stride of 2-convolutional layers and depth and thus reducing the feature sizes. The proposed model can prevent the challenges of overfitting via transfer learning based on CNN pre-trained networks in comparison to training with random initial weights.

For better classification results, SVM and error-correcting output codes (ECOC) known as ECOC-SVM were applied, which is based on the multiclass. The ECOC method uses class prior probabilities to minimize misclassification costs. This approach produces $\frac{K(K-1)}{2}$ by using one-vs-one coding, binary SVM model's architecture, where K is the number of a distinct class (for the face palsy detection task, there are five severity classes (levels) from I (mild) to V (complete paralysis) that can be used to evaluate the palsy severity. Comparing the ECOC model to other multiclass models can help it perform better overall; however, if necessary, it can be quickly simplified to a binary classifier. In instances where there is a clear margin for the division of classes, the SVM classifier performs well. Furthermore, it has been demonstrated to work effectively with extreme dimensional data, especially when the number is more than the dimensions of the sample.

The ECOC-SVM classifier was trained by using the activations from the FC layer containing 1000 weights. The choice of the multiclass ECOC-SVM classifier anticipated that, in the majority of cases, palsy images with a palsy severity score could be applied. In this instance, it is similar to a linear kernel SVM classifier because it utilizes such binary-labeled YFP dataset images as 'normal' and 'palsy'.

4.4.2. Image Color Histogram Equalization model[2]

The data augmentation approach implemented was based on *Image Quality Reduction* (IQR) using four variants of photometric transformation methods. This model was applied in the analysis of identifying cassava leaf diseases, and the different phases of the proposed structure are schematically illustrated in Fig. 4.9 as follows. The different types of photometric methods applied are majorly based on histogram transform, and our choice of these photometric data augmentation methods was based on two modules which are orthogonal Chebyshev polynomial functions and IQRM.

1. **Image dataset augmentation:** A unique parameterized image histogram transformation technique has been applied, and augmented images for data transformation tasks have been generated. The trainable color space and the training of the network were increased to identify significant color value distribution-based features, which are more resilient to image blurring, by including images with an enhanced color value distribution. Additionally, biological neural networks have been shown to implement a transform similar to histogram equalization, which provides a strong basis for applying

---

[2] The Material/Content presented in this section has been published in [60]

this ANN technique for image identification. Fig. 4.9 describes a conceptual entity diagram of the proposed method for cassava disease detection. Fig. 4.10 represents the flow diagram of the steps involved in the cassava disease detection model.

The proposed method is centered around the convolution of orthogonal functions with PDFs based on the color space values. A collection of fundamental functions known as orthogonal functions is available for describing any function in the function space.

Based on the findings, there are no prior applications of these algorithms for histogram equalization or data augmentation to solve classification and image processing issues. Chebyshev polynomials are unique variations of orthogonal ultraspherical polynomials. Type I Chebyshev polynomials $T_n(x)$ were created with an integer $n \geq 0$, and it is expressed as $T_n(\cos\theta) = \cos(n\theta)$. Therefore, Equations 4.4–46 show a few Chebyshev polynomials by using the Type 1 rule:

$$T_0(x) = 1, \tag{4.4}$$
$$T_1(x) = x, \tag{4.5}$$
$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{4.6}$$

Where $Tn(x)$ is the polynomial of degree n, and the formulas above show an enthralling pattern in Tn(x) and the sample representation is presented in Fig. 4.11 (right). By scaling the resultant function so that its total equals 1, the transformation is accomplished by using the product of the Chebyshev polynomial by the applicable PDF value. Fig. 4.11 (right) was created based on a mixture of Chebyshev orthogonal functions from orders 2 and 5 applied to each of the three-color channels (Red, Green, and Blue). Figure 4.11 (left)illustrate the image transformation using varied Chebyshev orthogonal functions and thus produces synthetic PDF transformation images on Fig. 4.11 (right) showing 24 distinct fake images, where the final cell is left empty. An instance of the transformation sample of the Probability Density Function (PDF) of the color value is represented in Fig. 4.12 with the original image, the PDF channel, and the transformation function for a synthetic image. The highlighted DAM methods were based on subsets of image corruptions or deformation types [251].
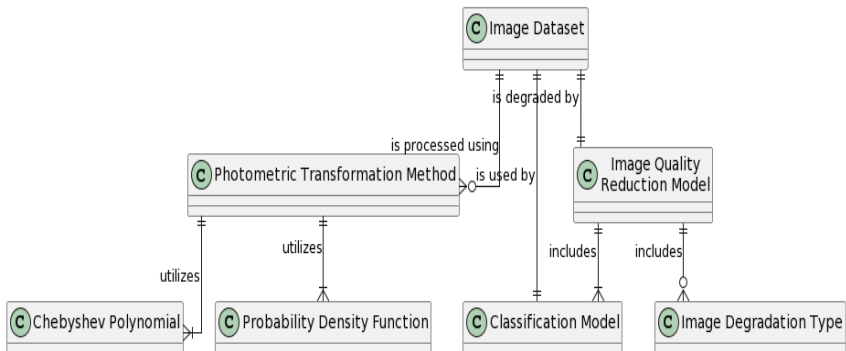
Fig. 4. 9. Entity diagram of the proposed method for cassava disease identification
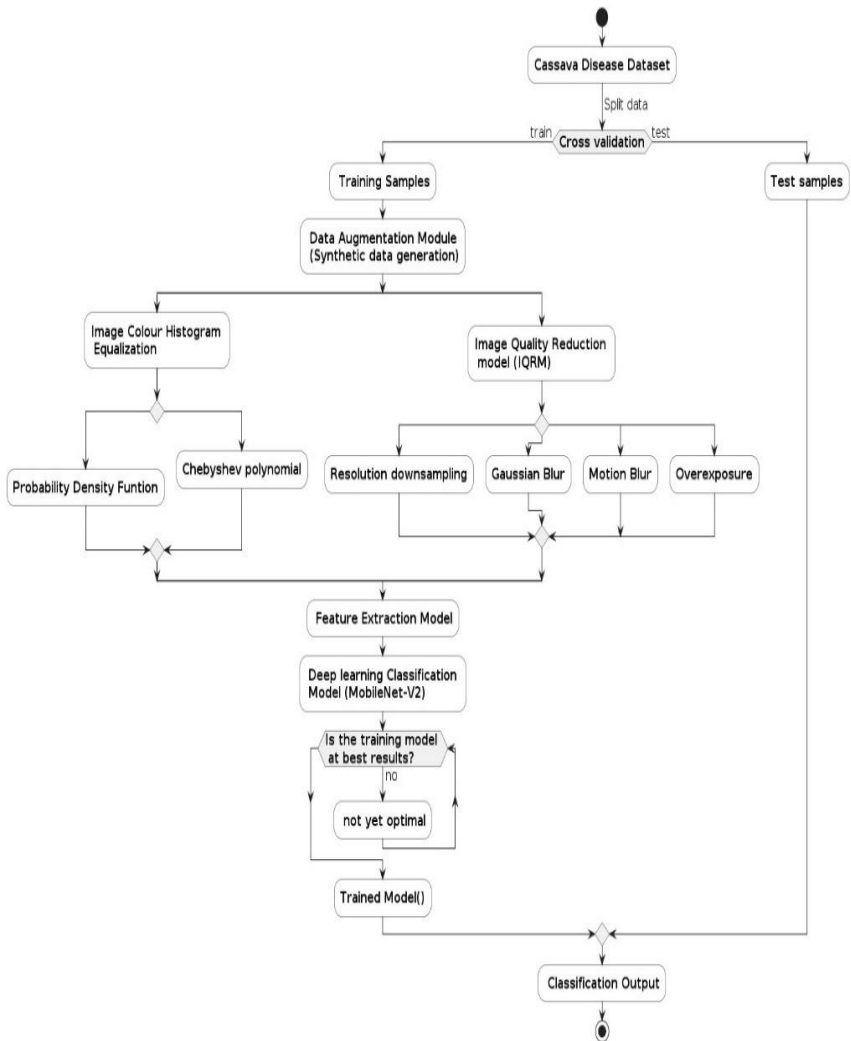
Fig. 4. 10. Activity UML-diagram of the cassava disease detection showing the Image Color Histogram Equalization and IQRM model
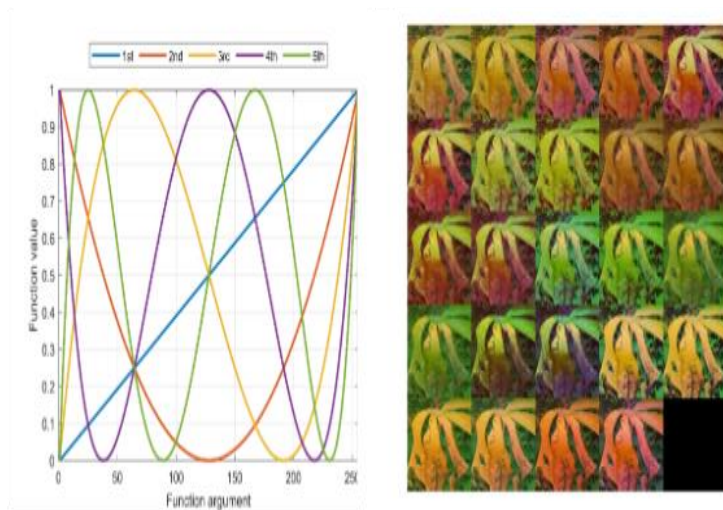
Fig. 4. 12. (left) Sample representation of orthogonal Chebyshev polynomial functions; (right) Different transformation of an image sample using different combinations of the probability density function (PDF)
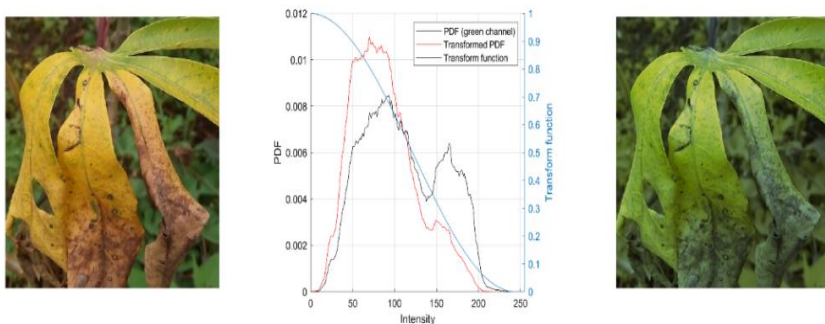
Fig. 4. 11. Transformation sample of probability density function (PDF) of the color value: original image (left); PDF of green channel (middle); transformation function and transformed pdf, and synthetic image (right)
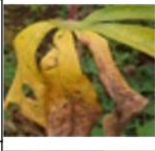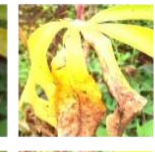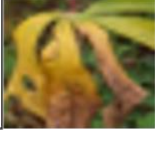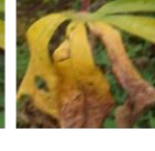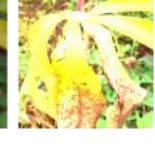
2. **Image quality reduction model (IQRM):** It has been observed that the modern CNN is susceptible to novel styles of known classes and unknown instances of recognized objects. Given this, even when utilizing cutting-edge CNN architectures like VGG-19 or ResNet50, the results of object detection with poor-quality images may be significantly inferior. The application of image quality reduction assumes that the general problem associated with real-life data is image degradation, which can result from the use of inferior devices by novice users in remote rural areas. The low quality of the mobile gadgets, motion blur, and excessive exposure caused by the camera in real-world scenarios are therefore predicted to have an impact on the images. As a result, the data augmentation approaches used were designed to degrade the quality of the test images, and, as a result, were

61

able to extract the lower quality of images as summarized in the below presented four transformation methods:

    a. *Resolution-down-sampling*: The raw image was down-sampled to the following pixels $32 \times 32; 64 \times 64; 96 \times 96; 128 \times 128; 160 \times 160; and 192 \times 192$ pixel and up-scaled to 224×224 pixel, as the final required input size of the neural network.

    b. *Gaussian-blur*: A Gaussian filter with a standard deviation of 1, 1.5, 2, 2.5, and 3 was used to convolve the raw input images.

    c. *Motion-blur*: This happens when a camera moves while taking an image. An evenly distributed random number was employed to determine the motion angle from a range of (0,360) angles, and the values of 10, 15, 20, 25, and 30 pixels were used as motion parameters to estimate the camera's linear motion.

    d. *Overexposure*: The amplitudes of the image channels were increased by a factor of 1.25 (1.5, 1.75, 2, and 2.5), thus replicating the overexposed photo conditions.

The methodologies adopted for this research were designed by using source [251], which implemented 15 types of image degradation and implemented five levels with each of those. A portion of these kinds of deformation was employed as certain original visual distortions, like frosted glass, are inherent in real-life situations. Tbl. 4.2 shows specific instances of samples of IQRM image modifications. The degrees of quality degradation are indicated as low, medium, and high-quality reductions.

Table 4. 2. IQRM model results for the four transformation methods used show the three levels of degradation applied L0–H2

3. **Classification Model:** For the automatic feature extraction and classification, a pre-trained *MobileNet-v2* model was adopted [65, 252]. MobileNet-v2 has been pre-trained with the *ImageNet* dataset and is based on an inverted residual framework where intermediate expansion layers perform depth-wise convolutions to filter features as a source of non-linearity, as depicted in Fig. 4.13. In addition, there are linear bottleneck layers which aid in preventing the loss of information resulting from nonlinearities. The network model performs better in the overall model complexity and accuracy than other real-time detectors [252]. The NN framework is relatively new, and it has already been successfully applied to plant disease recognition [253]. In other words, from low-level pixels to high-level forms, the neural network attempts to learn visual properties. Extraction of the intermediate-level visual feature was anticipated, which is crucial for classifying plant diseases by almost completely splitting the neural network in half. An FC layer, a softmax layer, and a classification layer with a cross-entropy loss function were added after the removal of the first 18 layers from the dataset, as illustrated in Fig. 4.13. The preliminary investigation did not show an improvement in the
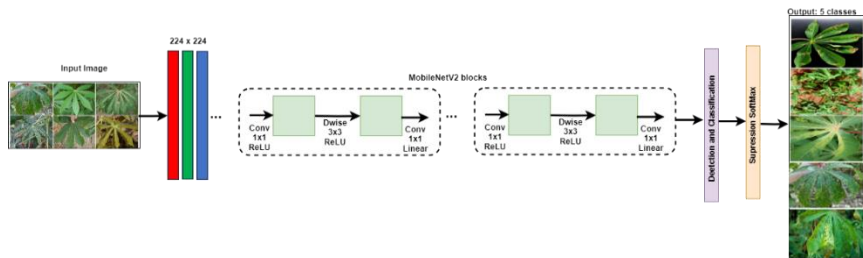
Fig. 4. 13. Baseline convolutional neural network framework

classification accuracy with extracted larger network models (with at least 18 layers). The following provides a more thorough explanation of the neural network's first 18 layers: A preprocessing layer is required after the input layer, which requires $224 \times 224 \times 3$ pixel images. Batch-normalization and clipped *Rectified Linear Unit* (ReLU) layers are placed after a Conv1-layer with $32\ 3 \times 3 \times 3$ convolutions with [2 2] stride. Batch-normalization and clipped ReLU layers are again applied after the grouped-conv-layer, which consists of 32 groups of $3 \times 3 \times 1$ convolutions with [1 1] stride. The BN layer is then followed by a conv-layer with $16\ 1 \times 1 \times 32$ convolutions, and then a conv-layer with $96\ 1 \times 1 \times 16$ convolutions is followed by the BN layer once more and the clipped ReLu layer. BN and clipped ReLU layers are again placed after the grouped-conv-layer, which consists of 96 groups of $3 \times 3 \times 1$ convolutions with [2 2] stride. Following the BN layer is the final conv-layer with $24\ 1 \times 1 \times 96$ convolutions, as depicted in Tbl. 4.3.

Table 4. 3. Baseline CNN parameters: First-18 layers are part of MobileNetV2, while the final three layers are added to classify cassava disease into five groups

#	Type	Activations	Learnable
1	Input_1 224 × 224 × 3 images	224 × 224 × 3	-
2	Preprocessing Preprocessing for MobileNet-v2	224 × 224 × 3	-
3	Conv1 32 3 × 3 × 3 convolutions stride [2 2] and padding 'same'	112 × 112 × 32	Weights 3 × 3 × 3× 32 Bias 1 ×1× 32
4	Batch Normalization (bn_Conv1) Batch Norm	112 × 112 × 32	Offset 1 ×1× 32 Scale 1 ×1× 32
5	Conv1_relu Clipped ReLU with ceiling 6	112 × 112 × 32	-
6	Expanded_conv_depthwise 32 groups of 1 3 × 3 × 1 conv	112 × 112 × 32	Weights 3 × 3 × 3× 32; Bias
7	Expanded_conv_depthwise_BN. Batch Normalization with 32 channels	112 × 112 × 32	Offset 1 ×1× 32 Scale 1 ×1× 32
8	Expanded_conv_depthwise_relu. Clipped ReLU	112 × 112 × 32	-
9	Expanded _conv_project 16 1 ×1 × 32 conv stride[1 1]	112 × 112 × 16	Weights 3 × 3 × 3× 16
10	Expanded_conv_depthwise_BN. Batch Normalization with 16 channels	112 × 112 × 16	Offset 1 ×1× 16 Scale 1 ×1× 16
11	Block_1_expand 96 1 ×1 × 16 convolutions stride [1 1] and padding 'same'	112 × 112 × 96	Weights 1 × 1 × 1× 16 Bias 1 ×1× 16
12	Block_1_expand_BN Batch Normalization with 96 channels	112 × 112 × 96	Offset 1 ×1× 96 Scale 1 ×1× 96
13	Block_1_expand_relu Clipped ReLU with ceiling 6	56 × 56 × 96	-
14	Block_1_depthwise 96 groups of 1 3 × 3 × 1 convolution	56 × 56 × 96	Weights 3 × 3 × 1× 1× 96 (bias)
15	Block_1_depthwise_BN Batch Normalization with 96 channels	56 × 56 × 96	Offset 1 ×1× 96 Scale 1 ×1× 96
16	Block_1_depthwise_relu Clipped ReLU with ceiling 6	56 × 56 × 96	-
17	Block_1_project 24 1 ×1 × 96 convolutions stride [1 1] and padding 'same'	56 × 56 × 24	Weights 1 ×1 ×96× 24 Bias 1 ×1× 24
18	Block_1_project_BN Batch Normalization with 24 channels	56 × 56 × 24	Offset 1 ×1× 24 Scale 1 ×1× 24
19	Fully Connected layer (FCl-5)	1 × 1 × 5	Weights 5 × 75264 Bias 5 ×1
20	Softmax (SC)	1 × 1 × 5	-
21	Classification output CL (5 classes)	-	-

4.4.3. Proposed CovSMOTE augmentation technique for Skin melanoma detection[3]

The skin image classification framework presented uses a variation of SMOTE for oversampling and image augmentation. Furthermore, to reduced dimensionality of image embedding manifolds, and image classification, a deep learning architecture based on SqueezeNet model was implemented [26]. The three phases, as depicted in detail in Fig. 4.14 and Fig. 4.15 show the structure described in the subsections below:
1. Pre-processing phase
2. Data Augmentation phase;
3. Classification Model
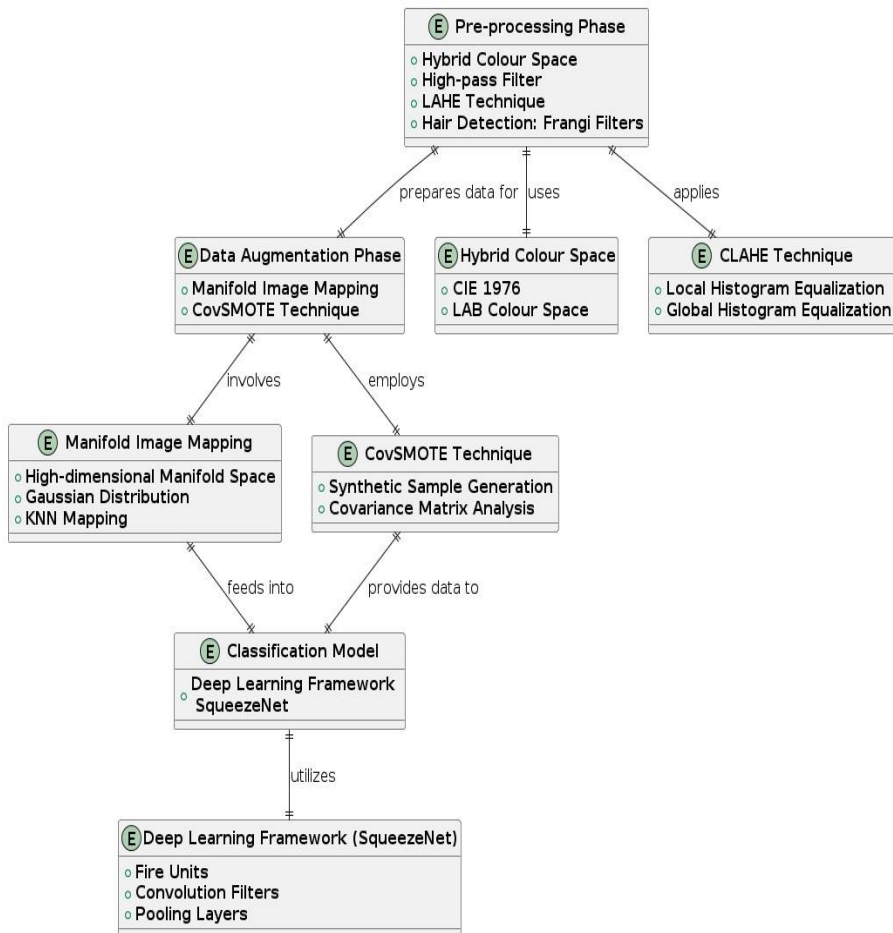
Fig. 4. 14. Entity diagram of skin melanoma detection

---

[3] The Material/Content presented in this section has been published in [271]
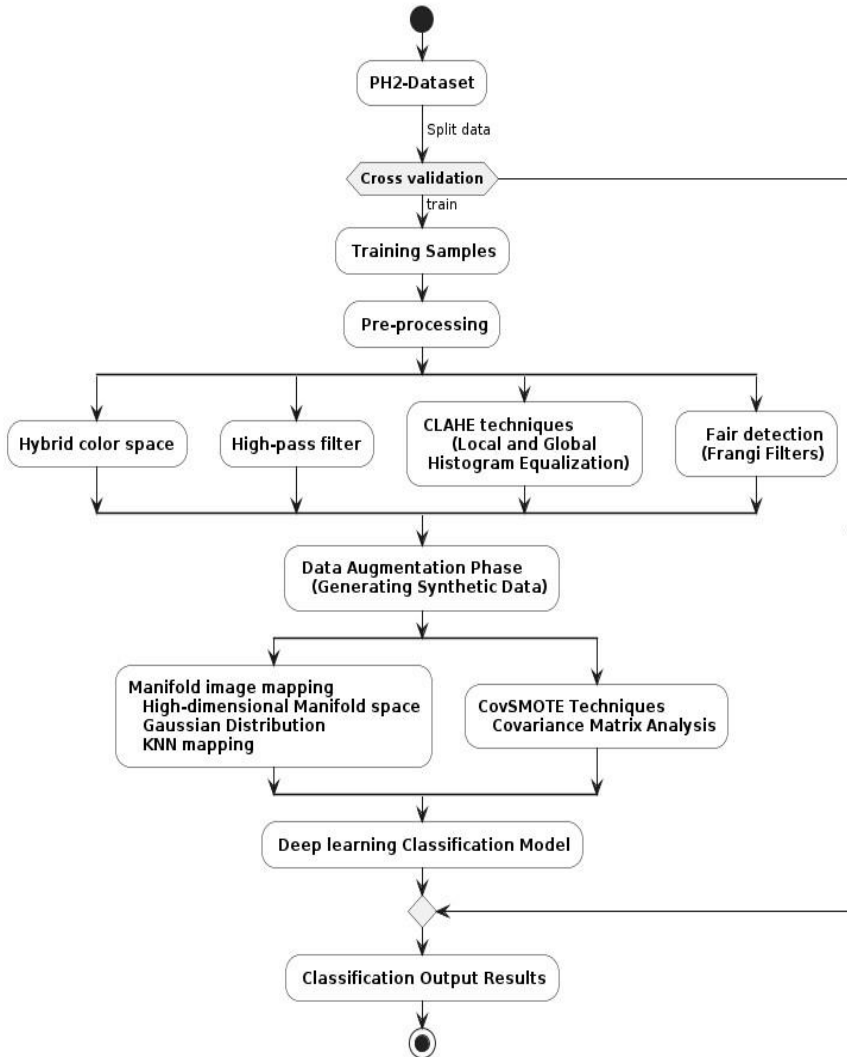
Fig. 4. 15. Activity UML-diagram of the skin melanoma detection framework

1. **Pre-processing:** Data pre-processing is essential for assuring data preparation and cleaning. The importance of mitigating errors during the training of classifiers is necessary so that a convolutional network can differentiate between the healthy skin class from low contrast, skin artefacts (namely lack of frames, and hairs), irregular borders in skin lesions, etc. In this phase, four pre-processing steps were adopted, as discussed below. The first phase, i.e., the hybrid color space, was proposed by integrating several color elements to increase their efficiency and decrease the correlation dependency between them. Standard illuminant D65, which has a color temperature of 6504 K and simulates noonday sunlight with values of 0.9504, 1.0000, and 1.0888, was used. In addition, RGB images were converted to

66

CIE 1976. Hering's theory was implemented to generate the $L * a * b$ (Lab) color space so that to enable the evaluation of a 'small' color. The theory was applied to demonstrate the hue component of the skin images Red, Green, Blue, and Yellow (R, G, B, Y).

i. The LAB color space has a three-dimensional color system where lightness (L*), and color (a* and b*). L* is the luminance shown in Equation 3.12, and values 0 and 100 are used to denote pure black and pure white, respectively. Equations 4.7 to 4.9 are used to indicate a pixel's lightness, whereas the a* and b* axes denote redness/greenness and yellowness/blueness as seen in Equations 4.10, respectively:

$$L* = \begin{cases} 116\left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} - 16 & for\ \frac{Y}{Y_w} > 0.008856 \\ 903.3\left(\frac{Y}{Y_w}\right) & for\ \frac{Y}{Y_w} \leq 0.008856 \end{cases} \tag{4.7}$$

$$a* = 500\left[f\left(\frac{X}{X_w}\right) - f\left(\frac{Y}{Y_w}\right)\right];\ b* = 200\left[f\left(\frac{Y}{Y_w}\right) - f\left(\frac{Z}{Z_w}\right)\right] \tag{4.8}$$

$$f(\varsigma) = \begin{cases} (\varsigma)^{\frac{1}{3}} & for\ \varsigma \geq 0.008856 \\ 7.787(\varsigma) + \frac{16}{116} & for\ \varsigma \leq 0.008856 \end{cases} \tag{4.9}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412 & 0.357 & 0.180 \\ 0.212 & 0.715 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.10}$$

Here $X,\ Y,\ Z$ are CIE 1931 tristimulus values of the image calculated. In addition, a high-pass filter was used to reduce the brightness heterogeneity and sharpen the borders of the skin lesions. As shown in Equation 4.11, the high-pass filter is generated by subtracting the weighted values of the Gaussian blur from the raw image.

$$I_f = I -_w G_\sigma(I) \tag{4.11}$$

where I is original image, if is a high-pass filtered image, G_σ is a 2-dimensional Gaussian kernel, $w \subset weighting\ [0,1]$ is an aspect that regulates the extent of image improvement, and σ is the Gaussian window width.

ii. Enhancing the image contrast, presents the *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) technique. The CLAHE algorithm efficiently minimizes the global noise by maximizing the contrast enhancement on local image data. In addition, every distinct contextual region is subjected to standard histogram equalization; thus, clipping and median filtering are applied. By assessing local and global histogram equalization, the CLAHE algorithm is capable of avoiding gray-level peaks. Fig.4.16 illustrates the outcomes of image enhancement to define two CLAHE parameters.
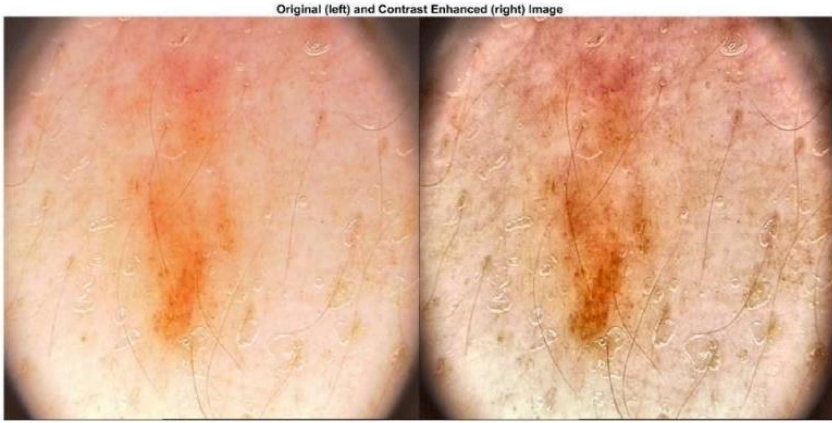
Fig. 4. 16. Sample of an enhanced image from PH2 dataset: (a) raw image; (b) transformed image

iii. Hair is detected with Frangi filters [254]. The second-order Gaussian derivative of image I at position (x, y) and scale σ yields the Hessian matrix H of the image as in Equation 4.12, while the Frangi function is mathematically represented in Equation 4.13:

$$H = \frac{d^2 I_\sigma}{dx^2} = I(x, y) \frac{d^2 G(\sigma, x, y)}{dxdy} \tag{4.12}$$

$$F(\sigma, x, y) = e^{\left(-\frac{R^2 \beta}{2\beta^2}\right)} \left(1 - e^{\left(-\frac{S^2}{2e^2}\right)}\right) \tag{4.13}$$

Where β and c are real positive and user-defined parameters; $R_B = \frac{\lambda_1}{\lambda_2}$, and λ_1 and λ_2 are the sorted eigenvalues of the Hessian matrix of a 2-D image, $|\lambda_1| \leq |\lambda_2|$ with β and c regulated to 0.5 and 15, respectively. A Frangi filter function was applied to remove hair from the background with the scale span set at 0.01 and 7.01, with each scale's step increasing by 0.5. The sample of the hair removal outcome is depicted in Fig. 4.17.

(a) (b) (c)
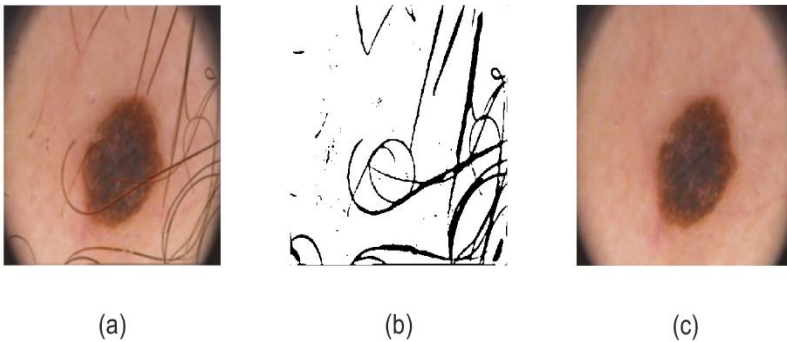
Fig. 4. 17. Example of the results of hair removal: (a) raw image; (b) the same image with hairs removed; (c) final image after hair removal

68

2. **Lower-dimensional manifold image mapping:** The skin images were mapped to minimize the dimensional manifold space by using the following steps: A high-dimensional manifold space was developed by using a differential manifold where the local space follows the Euclidean-style space for all skin image pixels, as represented in Algorithm 2 depicted in Tbl. 4.4.

Table 4. 4. Algorithm for Image manifold space mapping

Algorithm 2. Image manifold space mapping
Input: Image data X (locally Euclidean space)
d (dimensional topological manifold)
Start:
For $\forall$ point in X $= (x_1, x_2, \dots, x_d)$
$R^d = (x \in X)$
$\vee \leftarrow C^k$ differential structure
$V = \{(U_\alpha, \emptyset_\alpha)\}_{\alpha \in I}$ is the d-dimensional $C^k\ atlas\ (0 \le k \le \infty, a)$
The manifold M created by $(U_\alpha : \alpha \in I)$,
Mapping of $(\emptyset_\alpha : U_\alpha\ \text{to}\ \emptyset_\alpha(U_\alpha); and\ R^d)$
If $(U_\alpha \cap U_\beta \ne \emptyset, \&\&$ double mapping
$\emptyset_\beta^\alpha = \emptyset_\beta°\emptyset_\beta^{-1} : \emptyset_\alpha(U_\alpha \cap U_\beta) \rightarrow \emptyset_\beta(U_\alpha \cap U_\beta)\ )$
then $(U_\alpha, \varphi_\alpha)$ is compatible with (U_β, φ_β).
end.

The mapping in Algorithm 2 shows the class of differentiability C^k with $\emptyset_\beta^\alpha$ illustrating the differential mapping where the chart from $(U_\alpha, \varphi_\alpha)$ to chart (U_β, φ_β) is shown as a transformation of coordinates x with d-dimensional manifold. However, to effectively minimize the manifold space dimension, the Gaussian distribution approach was employed to create a low-dimensional space. The n-dimensional manifold space X of the skin image data can be defined by Equations 4.14 to 4.16. In addition, to complete the operation of skin classification, a distribution characteristic of data (T) is specified within the high-dimensional manifold space, and the distribution differences are decreased. By using a discrete distribution, random projection is applied to a plane of image data points.

$$x_i, x_j, \qquad x_i, x_j \in X, \qquad 1 \le i, j \le T \tag{4.14}$$

$$x_i = (x_i^{(1)}, x_i^{(2)}, x_i^{(3)})^T \tag{4.15}$$

$$x_j = (x_j^{(1)}, x_j^{(2)}, x_j^{(3)})^T \tag{4.16}$$

Here, the image data-position x_i indicates the centre of the Gaussian distribution of the points. The Euclidean distance identifies the distance between the projection position in a 2D skin image and the surrounding neighboring position with space L_p of x_i and x_j. The aforementioned processes can be used to develop KNN the skin pixels, and the local features

of the multiple data points are represented by a resulting KNN map. From Eq. 4.17, the probability $P_{j|i}$ is inversely proportional to the x_j distance:

$$P_{j|i} = \frac{exp\ (-\|x_i-x_j\|^2/2\sigma_i^2)}{\sum_{k\neq i} exp\ (-\|x_i-x_k\|^2/2\sigma_i^2)} \tag{4.17}$$

where probability $P_{j|i}$ is used by point x_i to choose point x_j as its closest neighbor. x_i is the center point T-distribution, while parameter σ^2 is the variance of the distribution which is used to demonstrate the local connectivity of the image position in the low-dimensional manifold to avoid the crowding issue. To successfully maintain the symmetry of the two-likelihood distributions in the two-dimensional space, a consistent symmetrical distance function is presented in Equations 4.18 and 4.19:

$$q_{i|j} = \frac{(1+\|y_i-y_j\|^2)^{-1}}{\sum_{k\neq i} e(1+\|y_k-y_l\|^2)^{-1})} \tag{4.18}$$

$$p_{ij} = \frac{p_{j|i}+q_{i|j}}{2n} \tag{4.19}$$

Here, n is the data points, $q_{i|j}$ is the likelihood distribution of the position of an image in the two-dimensional space, while y_i and y_j are two positions in a low-dimensional space. To tackle the overfitting problem caused by small data samples, a certain amount of confusion is also used, and the entropy value changes by modifying the level of confusion as the modification to the confusion level is proportional to the entropy. By aggregating the data position of the skin image in a two-dimensional space of positive and negative samples, we further develop the objective function, as represented in Equation 4.20 which aims to determine the weight between points.

$$P(e_{ij} = 1) = f(\|y_i - y_j\|^2) \tag{4.20}$$

Here, $P(e_{ij} = 1)$ depicts the likelihood between the binary edges of two points. However, $P(e_{ij} = 1)$ increases as the distances between y_i and y_j decreases. For the optimization procedure, it is necessary in the weighted edge of KNN graphs to expand the likelihood of the positive samples and minimize the probability of the negative samples with the initial transformation of Equation 4.21:

$$O = \sum_{(i,j)\in E} w_{ij} P(e_{ij} = 1) + \sum_{K=1}^{M} \sum_{jk} \sim_{p_n} (j)\gamma \log\left(1 - p(e_{ij} = 1)\right) \tag{4.21}$$

Here, γ is the negative sample weight, and thus, the resource cost is high by the negative sample $\underline{E}$, and this makes it harder to train the model by using the gradient descent swiftly. As a result, when selected carefully, a negative sampling procedure and the utilization of the noise distribution produced a negative sample $P_n(j)$ using a set of skin image points that were randomly selected.

3. **Proposed CovSMOTE Augmentation:** For data augmentation, a modified *Synthetic Minority Over-sampling Technique* (SMOTE) method was adopted [153]. The application of SMOTE generates data in feature space and produces synthetic samples by over-sampling the minority class. SMOTE focuses on recognizing k-minority class neighbors which are close to the minority class by arbitrarily choosing a point between the raw sample and the neighboring sample. The Cov-SMOTE [155] approach was applied, which is a modified SMOTE which adopts the Covariance Matrix to identify the dependency relationship between the attributes. From the estimate of the covariance matrix, new or surrogate instances were generated to obtain balancing between both classes (minority and majority). This process is terminated immediately as soon as there is a balance between the two classes. The algorithm for the CovSMOTE method is represented in Tbl. 4.5.

Table 4. 5. Algorithm for the CovSMOTE approach

Algorithm 3: Covariant SMOTE

Input: Dataset X, Number of majority class samples |M+|
            Number of minority class samples |M-|
Output: Balanced dataset for minority class sample
Process:
    Step 1: Load dataset X;
    Step 2: Compute the imbalance ratio IR $= \frac{|M+|}{|M-|} \geq 1.5$;
    Step 3: If IR ≥ 1.5 then
    Step 4:   Estimate Covariance matrix
                    $\text{Cov(X)} = \frac{1}{n-1} \sum_{ij}(X_i - X)(X_j - X);$
    Step 5:   Calculate the mean (average) values for each attribute in the dataset.
    Step 6: $\veebar$ pair of attributes (i, j) in the dataset: a range is determined
                    by $min - max$ value; While X is not in equilibrium do:
    Step 7:   Generate new instance y based on the covariance matrix;
    Step 8:   if range $\neq$ true then:
    Step 9:   Add y to X
    Step 10:   else
    Step 11:   For $i \leftarrow 0$ to Y_i do
    Step 12:   If $Y_i < \min Y_j$, then
    Step 13: $Y_i = \min Y_j$,
    Step 14:   else If $Y_i < \max Y_j$, then
    Step 15: $Y_i = \max Y_j$,
    Step 16:   else
    Step 17:   Continue;
                                    End
                                End
                            End
                        Else
    Step 18:   Return X_i
    Step 19:   End

The flow diagram in Fig. 4.18 presents the steps involved in generating the synthetic samples from the covariance within the interval of each attribute. Fig. 4.19 and Fig. 4.20 show the manifold space outcome of the PH2 dataset. A similar technique was used to produce new samples of the minority class in the lower-dimensional manifold space. The transformation from the manifold space to the image space was important once new samples were produced. The polynomial regression approach was applied due to the non-linearity of manifold space, thus aiding to obtain the best transform within the synthetic neighbor. To obtain a mapping from the manifold space to the color space, a cubic polynomial regression was performed, and the results obtained (synthetic samples images) are presented in Fig. 4.21.
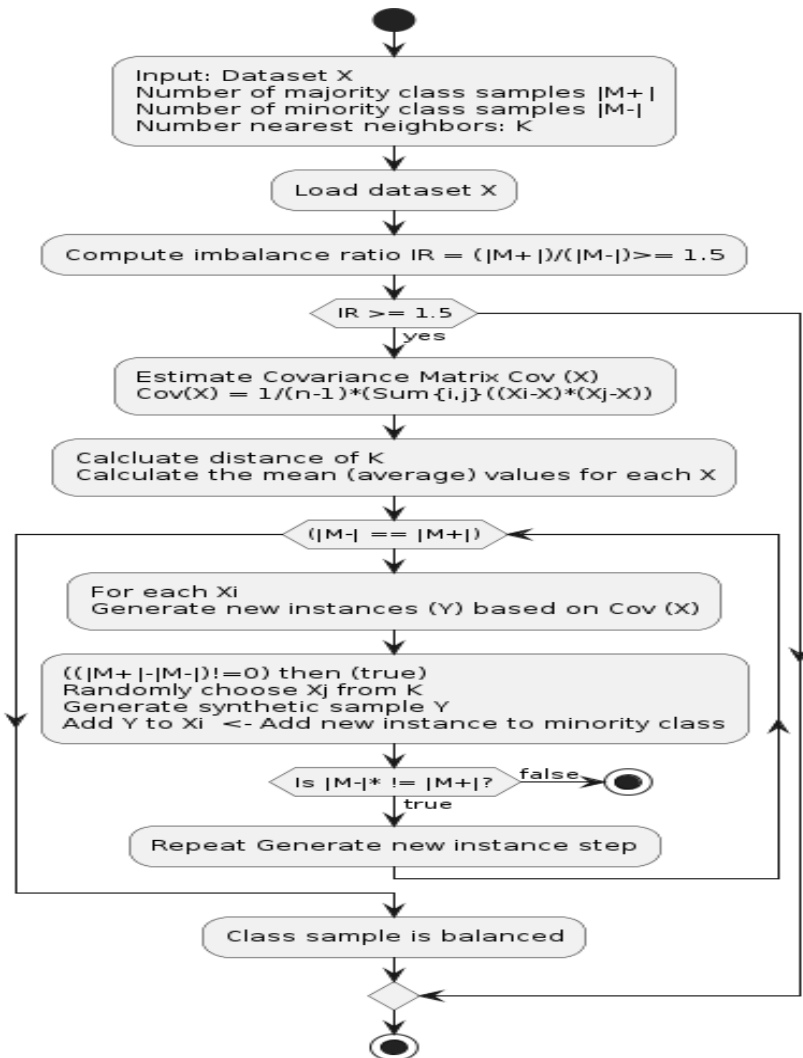
Fig. 4. 18. Flow diagram of the CovSMOTE method
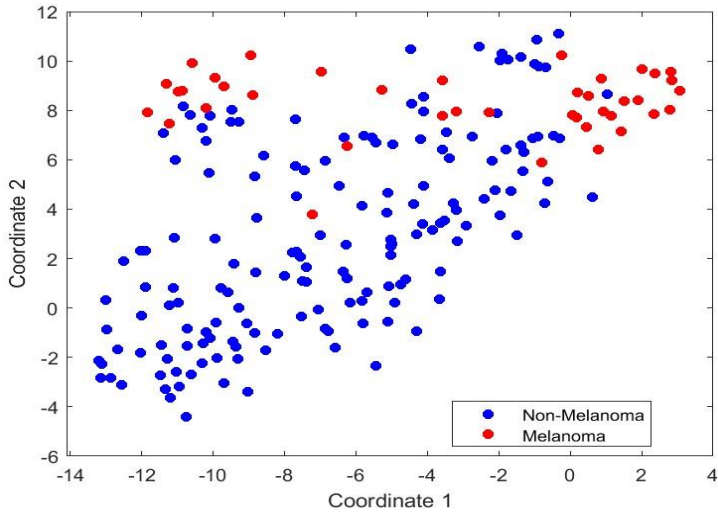
Fig. 4. 19. Visualization of PH2 dataset classes
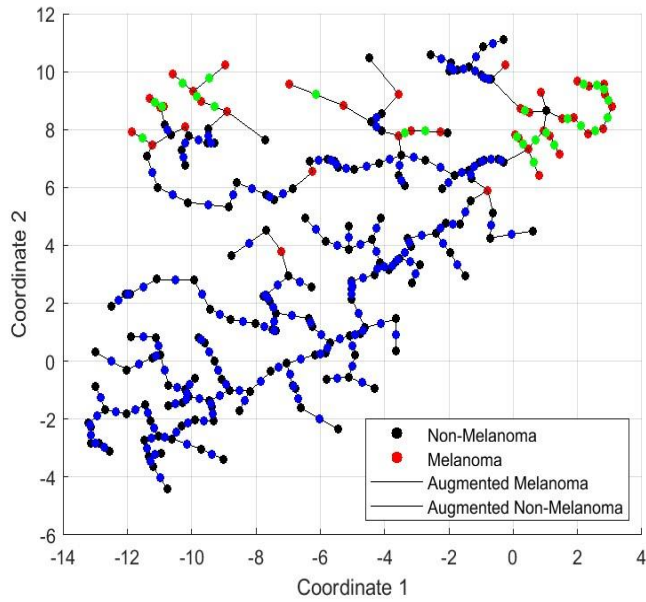
Fig. 4. 20. Representation of synthetic samples constructed; new instances are indicated with green dots
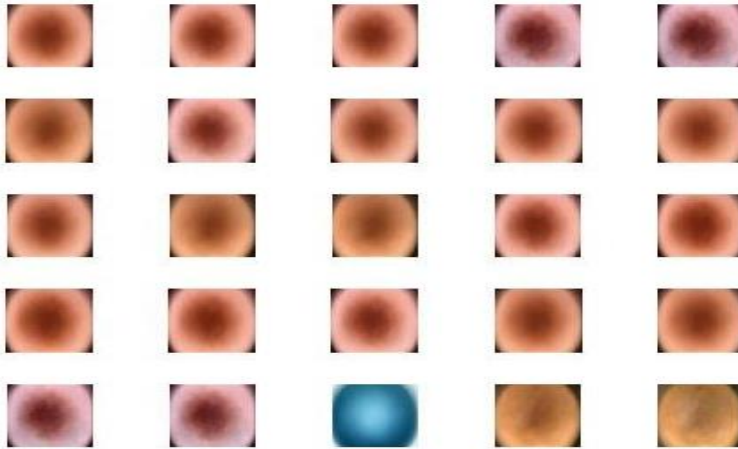
Fig. 4. 21. Results of images using CovSMOTE methods showing samples of generated augmented images

1. **Classification Model:** Deep learning techniques are crucial for successfully enhancing the classification accuracy and automating the removal of the necessary feature set. The identification of dermoscopic feature patterns in the skin lesion region, like blotches, streaks, blue-white veils, globules, and pigment networks, requires an efficient deep-learning algorithm. The *SqueezeNet* deep architecture was used to extract the relevant characteristics, whereas the choice of *SqueezeNet* is based on its ability to enable the operation of *AlexNet*-level outcomes with 50 times fewer parameters [250]. In the SqueezeNet framework, a few pooling levels and several fire units are stacked. To maintain the same size as the feature map, the fire unit must have both a squeeze layer and an expanded layer. The Squeeze layer parameters are decreased by using a 1×1 convolution kernel. Convolution filters with a mix of 1×1 and 3×3 is used in the expanded layer. The comprehensive results of the analysis shall be discussed in detail in the next chapter, and the outcome shows a DNN with outstanding performance while using few parameters.

4.5. Proposed Sound Augmentation Methods

4.5.1. Proposed Methodology[4]

The DeepShufNet framework is partitioned into data pre-processing, feature extraction, data augmentation, and classification modules, described in the subsections below. The conceptual diagram of the COVID-19 detection is presented in Fig. 4.22 and the framework for the proposed methodology DeepShufNet is illustrated in Fig. 4.23.

---

[4] The Material/Content presented in this section has been published in [181]
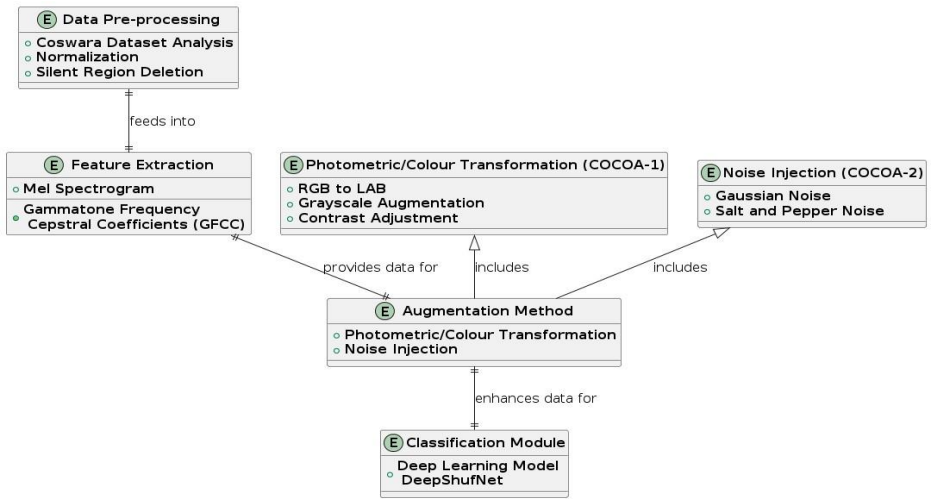
Fig. 4. 22. Entity Diagram of COVID-19 Detection
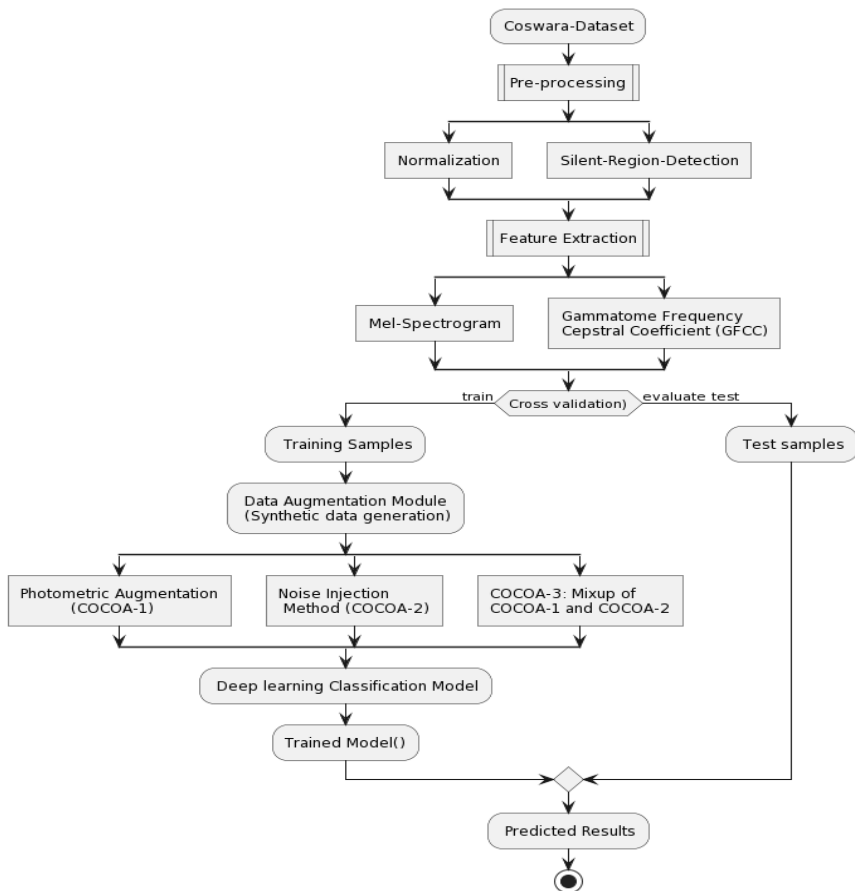
Fig. 4. 23. Activity UML-diagram of the proposed *DeepShufNet* method

4.5.2 Pre-processing

The COSWARA datasets consist of the irregular time span of signal for each audio recording file. To calculate the varying duration for each file, the mathematical equation $L = \left(\frac{N(Y)}{f(s)}\right)$ was applied with the sample length $N(Y)$ and the sampling frequency $f(s)$; every audio sample is 48KHz. The shortest and longest audio file lengths, according to the mathematical expression to compute (secs), are 4 seconds and 29 seconds, respectively. A simple yet straightforward pre-processing method was applied in terms of normalizing speech by using its maximum value and an amplitude maximum value of 1 so that to ensure that all pertinent features are recorded during the study. Furthermore, the silent region deletion approach was used to only use the voiced component of the signal speech and to remove the silent portion. The application of silent region elimination was already advanced in earlier investigations, thus inherently demonstrating its relevance in enhancing the system performance and cutting the processing time.

4.5.3. Feature extraction

For feature representation of the audio samples, two types of audio signal characteristics were used, which include Mel-spectrograms, and the *Gammatone Frequency Cepstral Coefficient* (GFCC) image, as highlighted below.

1. Mel Spectrogram: This is the time-frequency input representation that is most frequently utilized in sound classification tasks. In comparison to alternative representation structures, such as *Short-Time Fourier Transform* (STFT), this input representation has consistently demonstrated its usefulness and significance. The Mel spectrogram function via MATLAB toolboxes was used to convert COSWARA audio recordings into spectrograms. FFT window using a frequency range close to $2.0e^{-4}$ Hz, shows the creation of Mel spectrograms using audio recordings with an average length of 10 to 25 seconds. In addition, it may be inferred from the number of points surrounding the spaced times t and frequencies f that the power spectral energy density $P(f,t)$ for every audio file steadily increases for healthy samples in contrast to the samples of other class(es). A typical example of Mel Spectrogram images for each class is presented in Fig. 4.24.

2. *Gammatone Frequency Cepstral Coefficients* (GFCC): The use of Gammatone filter banks was first created in [255]; it was based on the human auditory system as an overlapping band-pass filter. The voice signal is enlarged to the gammatone filter banks in the frequency domain during the GFCC feature extraction procedure. A representation of a frequency-time signal called a cochleagram is created by using the output of gamma-filter banks. Hence, Equation 3.22 can be used to empirically represent the impulse response for each gammatone filter.

$$g(t) = mt^{y-1}e^{-2\pi nt}\cos\left(2\pi f_c t + \emptyset\right) \tag{3.22}$$

Here, $g(t)$ is the gammatone filter in the time domain, y which is commonly set to a value less than 4, determines the constant m, which normally equals 1 for the gain and the sequence of the filters. f_c is the frequency of the carrier, n denotes the bandwidth, and $\emptyset$ is the carrier phase, but it is typically set to zero. Equation 3.23 is used to compute the value of the function's *equivalent rectangular bandwidth* (ERB) for the auditory filter. Fig. 4.25 depicts the instances of GFCC time-frequency images for every class in the COSWARA dataset.

$$ERB = 24.7(\frac{4.37 f_c}{1000} + 1) \qquad (3.23)$$
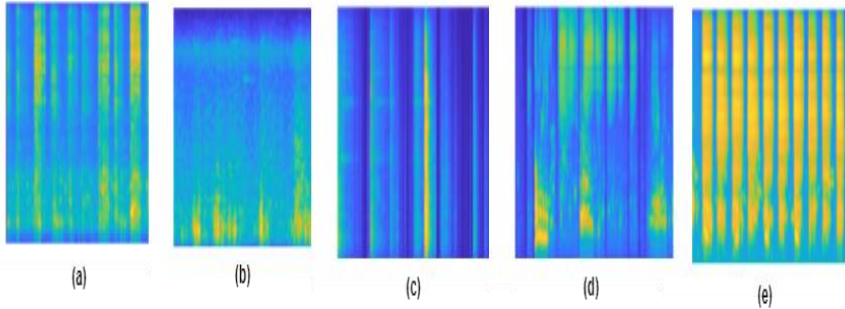
Fig. 4. 24. Instances of Mel Spectrogram images for five classes: (a) Positive_Asymptotic; (b) Positive_Mild; (c) Positive_Moderate; (d) Recovered_full; (e) Healthy
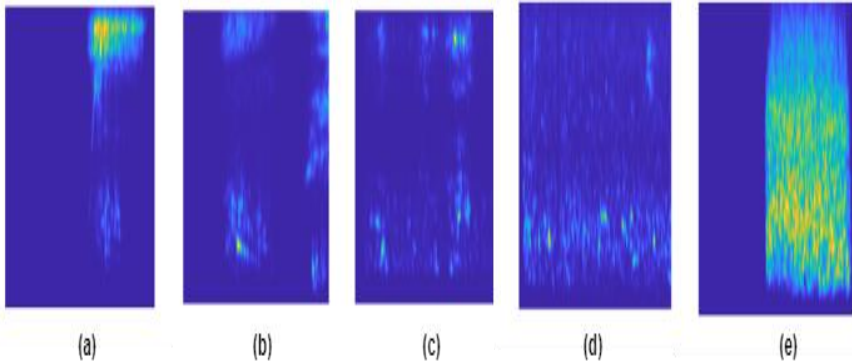
Fig. 4. 25. Instances of GFCC images for five classes: (a) Positive_Asymptotic; (b) Positive_Mild; (c) Positive_Moderate; (d) Recovered_full; (e) Healthy

4.5.4. Augmentation methods based on photometric and noise injection

Two kinds of data augmentation approaches have been proposed to enhance the minority class of the training images and thus create a new synthetic dataset. Photometric (or color transformation) and noise injection methods have both been employed to generate synthetic data and thus improve the training dataset. The description of the synthetic or augmented data called COswara-COvid-Augmented datasets coined as COCOA is described as:

1. The *Photometric* or *Color Transformation Method* (COCOA-1). There are three widely used color DAM models applied in the previous literature for improving data transformation. However, this study analyzed and investigated the rgb2lab and grayscale augmentation approaches. There are 256 shades of grey and a brightness range of 0 (black) to 1 (white); these transformer approaches are also known as monochrome. Among the other color transformation techniques used in augmenting the training data, these include contrast, rgb2lab, brightness, and rgb2gray.

2. The *Noise Injection Method* (COCOA-2). The most widely used intensity transform approach is the noise injection method (NIM). Any deterioration in the signal of an image which is often caused by an external disturbance is referred to as noise in images. Thus, the injection of noise in the image input data during training has been said to exert significant influence on the performance of learning models through improved learning and efficiency of the model [256]. To create synthetic datasets, different ranges of noise were adopted, like the salt-pepper noise and gaussian noise to each image in the dataset.

    i. *Gaussian Noise*: This is an analytical noise often referred to as Gaussian distribution with equal PDF in the normal distribution. The property of the Gaussian distribution is that it is characterized by the mean and the variance, σ^2. Considering an image $f(X)$ with coordinates (X, Y) with an additive noise of $n(x)$, hence, the Gaussian distribution pdf is computed as the mean of distribution (μ) by taking the average of the experimental values $m = \frac{1}{n}\sum_{i=1}^{n} x_i$, and m is the highest possibility estimate of μ using the mathematical expression in Equation 3.24.

    $$f_x(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3.24}$$

    σ is the standard deviation, and the effect of Gaussian noise is directly proportional to the value of σ. The probabilities for pepper and salt cases are assumed to be equal, and the total probability of the degradation of a pixel is referred to as the magnitude of the noise.

    ii. *Salt and Pepper Noise*: This is an impulse noise which is based on a vast range of processes such as sudden disturbance which could result in image degradation where just a few pixels are very noisy with an effect like sprinkling white and black dots often referred to as salt and pepper on the original image [257]. As an instance, an image (X) with number of pixels as (n) bits can be expressed as $X = \frac{1}{n}\sum_{i=1}^{n} N_i 2^i$, for $i = 1, 2, \dots$, N.

    Furthermore, also geometric methods have been applied, such as shear, zoom, and horizontal flip, to every image in the dataset to create an additional dataset. The COCOA-3 dataset is a combination of the COCOA-1 and COCOA-2 datasets.

4.5.5. Classification module (DeepShufNet)

The *DeepShufNet* model has been presented, trained and tested for the baseline and augmented dataset. Fig. 4.26 illustrates a lightweight deep CNN model called DeepShufNet which is based on pretrained ShuffleNet architecture. Previous studies showed that the pointwise Group Convolution theory which is represented in the ShuffleNet architecture has a light-weighted network which employs channel shuffle, recurrent modules, and allocates models over two processors. Additionally, the implementation of channel shuffling and pointwise group convolution has reduced the computing costs while maintaining the overall performance. 1D-CNN for binary classification was applied to train the audio recording which is a 1D- time series. The DeepShufNet is made up of an input layer with an image resolution of 224 x 224 x 3, several hidden layers, namely, convolutional, batch normalization, pooling layer, flatten layer, FC layer, and an output layer. However, each image was downscaled from its original size of 875 by 656 pixels to 224 by 224 pixels, which is sufficient for recognizing all target ranges. Additionally, the use of smaller inputs improves the computational speed, minimizes the parameter numbers, and ultimately reduces the likelihood of overfitting. The experiment employed the proposed DeepShuffleNet which has 172 layers overall and 1.4 million learnable parameters. The optimizer used in this experiment is the dropout layers of 0.5 to the hidden neurons so that to eliminate overfitting and to ensure the optimal model. The DeepShuffleNet architecture employs unique operations, such as grouped convolution, channel shuffle, and depth concatenation which drastically decrease the computing time and enhance the performance outcomes despite the higher number of layers. The training options applied are the *Adam* (adaptive momentum algorithm), 250 minibatch size for exploring and final training. The optimizer's learning rate is dependent on the warm start parameter, which ranges from $1e^{-4}$ to $1e^{-3}$, the total number of epochs, and the L2-Regularization parameter, which is $\lambda = 2 \times e^{-4}$.
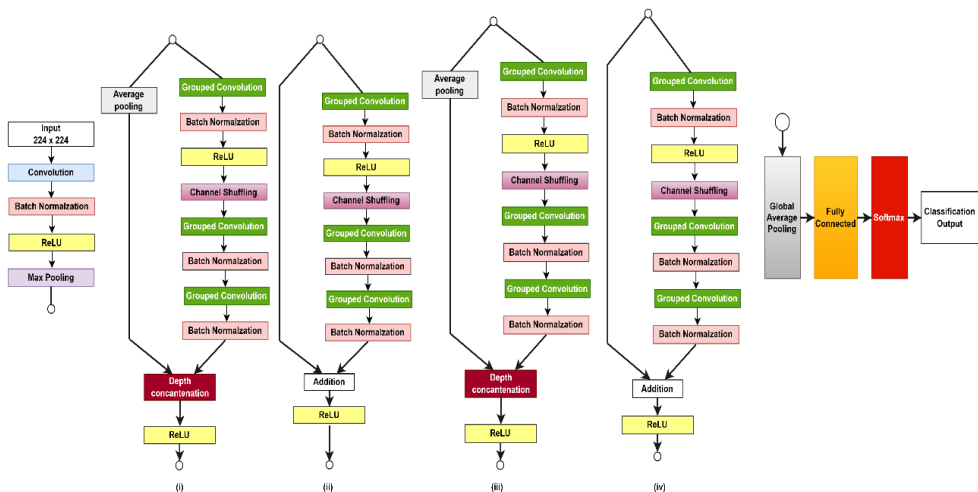
Fig. 4. 26. DeepShuffleNet Structure and Components

4.6. Performance Metrics

The results of the proposed DAM after training on the learning model have been evaluated. To calculate the efficacy of the proposed models, benchmarked metrics were adopted, such as TP – True positive, FP – False positive, TN – True Negative, FN – False negative, and the further optimization metrics used are: ACC –Accuracy, SEN – Sensitivity or Recall, SPEC – Specificity, PRE – Precision, F –score, Receiver Operating Characteristics Curve/Area Under Curve (ROC-AUC). A thorough breakdown of the performance metrics employed in this thesis is presented with mathematical expressions in Equations 3.25 – 3.29.

a. Accuracy (ACC): This is the likelihood that the classes in the entire sample of cases were correctly predicted.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \qquad (3.25)$$

b. Sensitivity/Recall: This is the likelihood of all true positive classes that are correctly predicted as positive.

$$Rec = \frac{TP}{TP+FN} \qquad (3.26)$$

c. Specificity: This is the likelihood of all true negative classes that are correctly predicted as negative.

$$Spec = \frac{TN}{TN+FP} \qquad (3.27)$$

d. Precision: This is the likelihood of classified positive classes that are correctly positive.

$$Pre = \frac{TP}{TP+FP} \qquad (3.28)$$

e. F1-Score: This is the weighted harmonic mean of precision and recall (sensitivity).

$$\text{F1-score} = 2 * \frac{PRE*REC}{PRE+REC} \qquad (3.29)$$

f. ROC: It is a standard technique for evaluating how effectively the classification and pattern-matching systems function. The trade-off between TPR and FPR can be shown by ROC. AUC determines the measure of classifier effectiveness in a single number; thus, AUC measures the reliability of the classification evaluation criteria. In addition, AUC is the likelihood that a positive instance will receive a higher score than a negative instance based on how the classifier rates positive cases to the negative instances, with the mathematical expression in Equations 3.30 to 3.32 where T is the varying parameter.

$$\text{True-Positive-Rate (TPR)} = \frac{TP}{TP+FN} \qquad (3.30)$$

$$\text{False-Positive-Rate (FPR)} = 1 - Specificity \qquad (3.31)$$

$$AUC = \int_{T}^{\infty} TPR.d\,(FPR) \qquad (3.32)$$

4.7 Summary of the Proposed Materials and Methods

In this chapter, a low computationally intensive method was proposed for the facial palsy dataset. This method is based on efficient data augmentation methods *Voronoi Decomposition Random Region Erasing* (VDRRE) approach. In addition, a lightweight and minimal complexity CNN model *SqueezeNet* was adopted to select high-level features and train the model. For cassava disease detection, data augmentation was implemented by using a color histogram model, and four image quality reduction methods were applied to overcome the issue of overfitting and poor generalization. Practical characteristics were taken from the CNN, and deep learning model hyperparameters were fine-tuned to improve cassava disease classification. An optimized and efficient CNN (MobileNetV2) model was introduced to detect cassava disease by using cassava leaf images. For the skin melanoma classification, a new data augmentation technique using the oversampling of a nonlinear lower-dimensional embedding manifold to generate augmented samples. A covariant *Synthetic Minority Oversampling Technique* (SMOTE) was presented to tackle the issue of limited data and class disparity. The proposed covSMOTE augmentation model was able to effectively create synthetic melanoma images and improve the performance of the training model. In COVID-19 detection by using sound data, color transformation and noise injection methods were applied to generate artificial data from sound data. The efficiency of training the proposed model by using the augmented dataset was analyzed by using two feature extraction methods which are the Mel spectrogram and GFCC on the proposed DeepShufNet model.

Another study used ensemble-based machine learning classifiers in the detection of COVID-19 by using blood test medical data. A custom classifier was introduced by using *Convolutional Neural Network* (CNN) models and 15 supervised machine learning algorithms. Finally, we adopted spline interpolation and *Piecewise Cubic Hermite Interpolating Polynomial* (Pchip) interpolation methods to create augmented data from known observations. However, the classification task was based on BiLSTM and conventional machine learning classifiers, such as Ensemble bagged tree, SVM, LR, DT and KNN. The significance of each method in various classification tasks shall be presented in Section 5.

5. EXPERIMENT AND RESULTS

In this chapter, various implementations and experimental outcomes of the proposed data augmentation (DAM) methods on the different databases outlined in the previous chapter shall be discussed in detail. The subsections in this chapter will address in detail the integrated DAM methods in different classification tasks.

5.1. Analysis of Face Palsy Detection

In this experiment, the VDRRE augmentation method discussed in Section 4.3 for face palsy detection was trained by using the deep learning (SqueezeNet) network. The CNN architecture used in this experiment was implemented in *MATLAB R2019a* running on *Windows 10* 64-bit *Intel Core i5* CPU and 8 GB RAM. The epochs were set to a maximum of 50, the learning rate was set to 0.00001, a fixed mini-batch of size 16 was used, and the *Stochastic Gradient Descent Momentum* (SGDM) optimizer was employed in the training process. To effectively train the model, the following additional parameters used for this experiment are summarized in Tbl. 5.1. Early stopping of training [58] was introduced to stop the training as soon as the validation loss has begun to rise, thereby enhancing the generalization ability of the model and avoidance of overfitting.

Table 5. 1. Configuration Model parameters for face palsy detection

Model Parameters	
Batch size	16
Optimizer	SGDM
'MaxEpochs'	50
Exponential decay rates	0.9
'WeightLearnRateFactor'	20
Epsilon parameters	0.999
Moment estimates	$10 \ x \ 10^{-8}$
'InitialLearnRate'	0.00001

5.1.1. Performance analysis results

The dataset was partitioned at random into five-fold cross-validation, and the training set comprises the four-fold making up 80% of the dataset, while the last fold i.e., 20% was used to test. Tbl. 5.2 summarizes the confusion matrices for the experimental results. Each experiment was carried out repeatedly, and the outcome of the model used the following metrics: recall, accuracy, F1-score, and precision. The confusion matrices display the results of numerous cross-validation folds combined. It should be noted that the rate of misclassification is minimal for the 1-shot and 2-shot-learning tests with image augmentation, which indicates high-performance levels which are reprinted in full in Tbl. 5.2. The $p-value$ metrics results for each data is summarized in Tbl. 5.3. The mean score with 95% confidence intervals is presented in Tbl. 5.4 which depicts the classification results for face palsy detection when compared to using the proposed image enhancement technique.

Table 5. 2. Confusion matrix with the influence of VDRRE on the test dataset

Original Image (No_Aug)				
Predicted/ Actual	Normal (Predicted)	Palsy (Predicted)	True Positive (%)	False Positive (%)
Normal (Actual)	67	10	87.0	13.0
Palsy (Actual)	23	211	90.2	9.8
1-Shot Learning + VDRRE				
Predicted/ Actual	Normal (Predicted)	Palsy (Predicted)	True Positive (%)	False Positive (%)
Normal (Actual)	82	2	97.6	2.4
Palsy (Actual)	8	219	96.5	3.5
2-Shot Learning + VDRRE				
Predicted/ Actual	Normal (Predicted)	Palsy (Predicted)	True Positive (%)	False Positive (%)
Normal (Actual)	83	1	98.8	1.2
Palsy (Actual)	7	220	96.9	3.1

The hybrid SqueezeNet/ECOC-SVM classifier was implemented in both 1-shot and 2-shot-learning using the proposed VDRRE approach. The integration of the trained 2-shot learning on VDRRE gave the best performance, by attaining precision, recall, F1-score, and accuracy, with scores of 99.35%, 99.74%, 99.54%, and 99.34%, respectively. The experimental outcome of the comparable 1-shot learning case was marginally poorer in terms of performance, with precision, F1-score, recall, and accuracy scoring values of 98.85%, 99.28%, 99.72%, and 99.07%, respectively. Nevertheless, the baseline result with no augmentation obtained F1-score, precision, recall, and accuracy yielded rates of 85.59%, 81.06%, 91.85%, and 78.62%, respectively. Both 1-shot and 2-shot learning with VDRRE augmentation, on the other hand, produced significantly improved results.

The t-distributed Stochastic Neighbor Embedding (t-SNE), which leverages PCA for reducing the dimensionality of features, was applied to efficiently visualize the potential of the SqueezeNet network to extrapolate the useful features. From Fig. 5.1, the nonlinear dimensionality-reducing technique, known as t-SNE, enables the visual representation of high-dimensional data as a two-dimensional map. The visualization shows that the two-dimensional embeddings of palsy face images form a cluster which is clearly distinguished from the two-dimensional embedding of normal face images. From Tbl. 5.4, it can be deduced that, in every way, the results were significantly improved when using either the random erasing augmentation or VDRRE (one-shot and two-shot learning). Thus, the most desirable results are shown in bold, and this shows the impact of the VDRRE model in enhancing the effectiveness of the classifiers with almost a 20% increment in the accuracy of the model with augmentation.

Table 5. 3. Comparison of statistical significance of the proposed best results for each learning scenario

Metrics	Statistics	No Aug	With Aug	
			1-Shot	2-Shot
Accuracy (%)	Average	78.62	99.07	**99.34**
	Minimum	63.59	97.10	98.87
	Maximum	91.16	99.7	99.80
	Standard deviation	7.89	0.72	0.34
	p-value	1.9189×10^{-7}	1.4669×10^{-6}	0.30332
Precision (%)	Average	81.06	98.85	**99.35**
	Minimum	72.61	95.45	98.66
	Maximum	90.31	99.77	99.66
	Standard deviation	6.29	1.40	0.33
	p-value	0.010096	0.0040585	0.87758
Recall (%)	Average	91.85	99.71	**99.74**
	Minimum	78.28	99.09	99.43
	Maximum	99.32	100	100
	Standard deviation	7.57	0.36	0.25
	p-value	7.7469×10^{-7}	3.3278×10^{-8}	0.29494
F1-Score (%)	Average	85.91	99.28	**99.54**
	Minimum	75.34	97.67	99.21
	Maximum	94.03	99.77	99.83
	Standard deviation	5.28	0.66	0.23
	p-value	2.7052×10^{-6}	1.9166×10^{-7}	0.252

Table 5. 4. Average-performing results of hybrid SqueezeNet/ECOC-SVM classifier for palsy detection

Methods	Average Classification results (95% CI)			
	Acc (%)	Recall (%)	Prec (%)	F1-Score (%)
No augmentation	78.62 (SD: 5.65)	91.85 (SD: 5.41)	81.06 (SD: 4.50)	85.59 (SD: 3.78)
Random-Erase augmentation	92.91 (SD: 1.12)	96.14 (SD: 0.83)	93.96 (SD: 1.87)	95.04 (SD: 1.42)
2-shot learning +VDRRE	99.35 (SD: 0.24)	99.74 (SD: 0.17)	99.35 (SD: 0.24)	99.54 (SD: 0.16)
1-shot learning + VDRRE	99.07 (SD: 0.60)	99.72 (SD:0.28)	98.85 (SD: 1.15)	99.28 (SD:0.55)
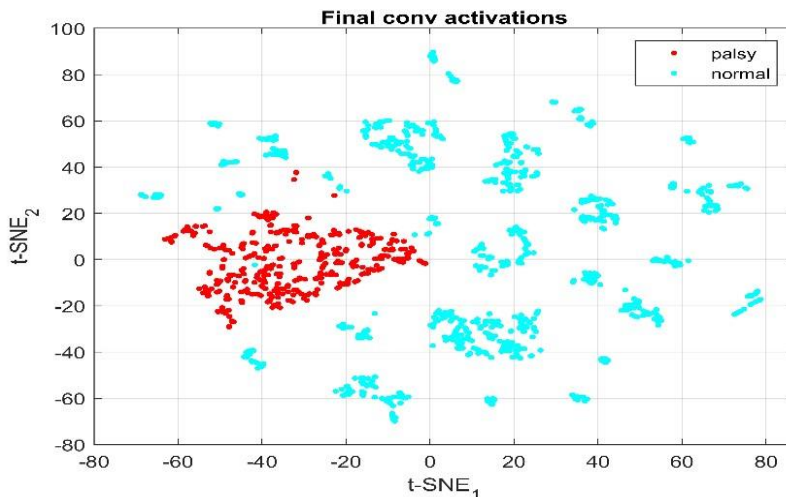
Fig. 5. 1. Data visualization by using t-SNE

Fig. 5.2 shows the ROC curve which was employed to further evaluate models in three scenarios (baseline results, 1-shot-learning with VDRRE, and 2-shot-learning with VDRRE). The AUC metric, which is represented using the ROC curves with has the following values: 0.7967 (95% CI = [0.7944, 0.7989]) for baseline, 0.9958 (95% CI = [0.9957, 0.9959]) for 1-shot learning, and 0.9956 (95% CI = [0.9955, 0.9957]) for 2-shot learning cases, respectively. In this instance, the classifier performance matrix values were bootstrapped to calculate the *Confidence Intervals* (CI) on the premise that they had a normal distribution. These findings demonstrate that the proposed VDRRE approach greatly outperforms 'baseline results' cases in both 1-shot and 2-shot cases.
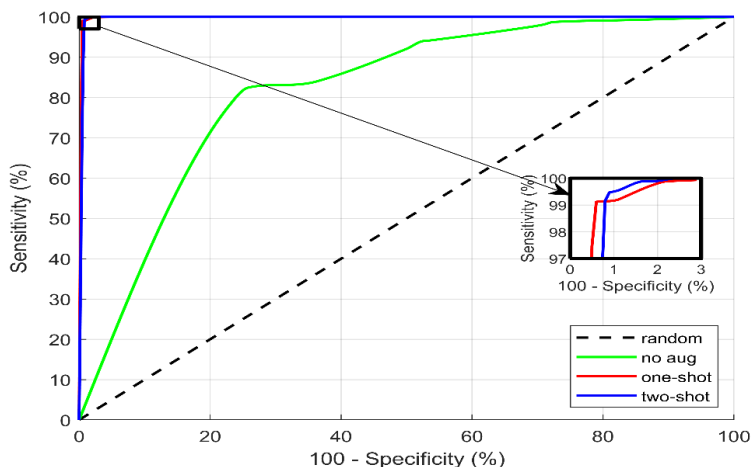
Fig. 5. 2. ROC curve displaying area under the curve for different augmentation test results
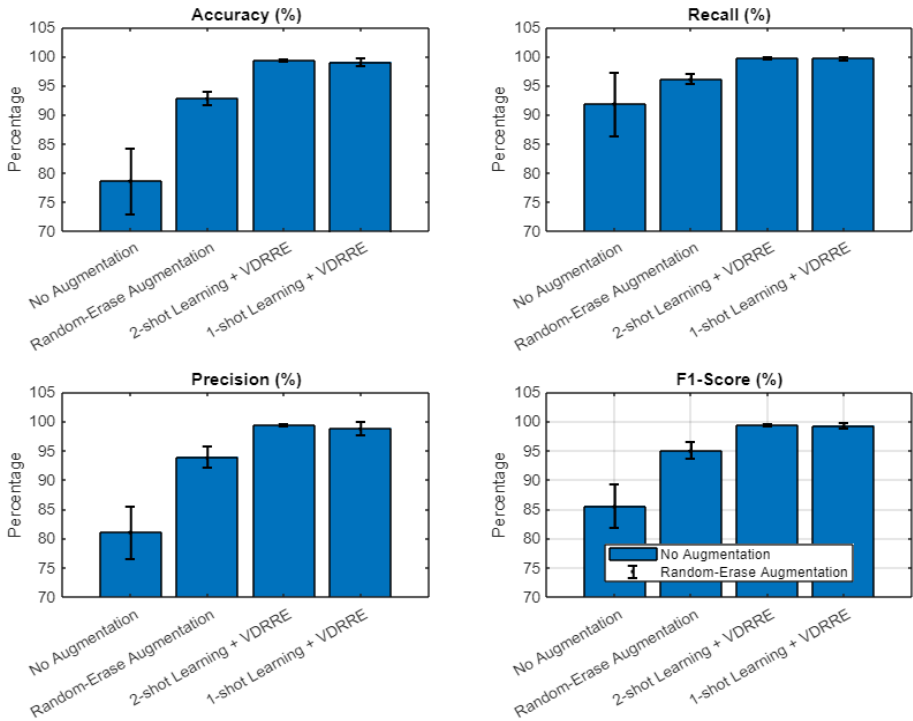
Fig. 5. 3. Overall outcome of 2-sampled t-test comparison for baseline, 1-shot and 2-shot-scenarios: (a) accuracy; (b) recall; (c) precision; (d) F1-score

Where: $*.p \leq 0.05$, $**.p \leq 0.01$, $***.p \leq 0.001$, $****.p \leq 0.0001$, n.s. – not significant $p > 0.05$.

To statistically analyze the findings, for equal means, a two-sample t-test was used. The outcome of the test provides a conclusion on the null hypothesis, which is that the data in the two compared samples are distributed normally, with equal means but unknown variances. The tests were run at the 5% threshold of significance. The findings shown in Fig. 5.3 demonstrate a substantial ($p \leq 0.001$) difference between the baseline and the 1-shot-learning case with VDRRE, in addition to the 2-shot-learning scenario with VDRRE. Nonetheless, there was little variation in performance between the 1-shot and 2-shot learning scenarios (i.e., the equal means hypothesis was not rejected).

5.1.2. Benchmark analysis with existing methods

To further validate our proposed method, a comparison of the experimental findings with the existing approaches sourced from [99, 246] on the same YFP dataset was conducted and reported in Tbl. 5.5. With a definite advance in recall, precision, and accuracy, the comparison demonstrates that the proposed model outperforms the previous studies. The application of the proposed VDRRE image augmentation

technique increased the image number accessible for NN training and ultimately contributed to the performance gain.

Regardless of the positive outcomes produced by the proposed VDRRE augmentation technique, the drawbacks of utilizing a limited dataset continue: (1) the diversity in expressing the degree of the facial palsy condition is lower in small datasets; (2) the risk of overfitting (leading to poor generalization of the DNN models) is still present. The binary nature of the experiment (i.e., normal vs. palsy), which does not distinguish between different degrees of facial palsy severity, may have an impact on the validity of our findings. This indicates that there may be more variation within the 'palsy' class than there is between the 'palsy' and 'normal' groups. The multiclass classifier ECOC-SVM which was employed in the last step of the workflow, however, makes the proposed methodology general and enables the use of various facial palsy datasets with severity-level class labels.

Table 5. 5. Comparison of performance results with the existing methods for face palsy detection by using the YFP dataset. Bold fonts convey the best values

Methodology		Performance results			References
Classifier	**Augmentation**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	
Deep Hierarchical Network	NA	91.20	-	-	[99]
Parallel Hierarchy CNN+ LSTM	GAN, translation, and rotation transformation	94.81	95.60	94.80	[258]
Our proposed model	No augmentation	89.25	95.43	89.13	Our paper
	Geometric & Color transformation	99.07	98.85	**99.72**	
	VDRRE (proposed)	**99.34**	**99.43**	99.35	

5.1.3. Conclusion of the section

This study has proposed a DL-based classification methodology for the identification and classification of facial palsy. The proposed methodology presented an innovative image enhancement technique which expanded random erasing enhancement with erratic regions created by using Voronoi tessellation. SqueezeNet deep neural network for automatic deep feature extraction was presented, and, finally, the workflow, based on a multi-class classifier, was developed. Therefore, the proposed approach can be used to evaluate facial palsy by utilizing a variety of datasets related to facial palsy, including multi-class datasets that contain face images labeled with the palsy severity level.

A few-shot learning methodology inspired by human intelligence was used to train our system to successfully distinguish palsy facial images within a small sample size. To create synthetic training sets for the scenarios of 1-shot and 2-shot learnings, the VDRRE image augmentation approach was put into practice. The proposed hybrid classifier uses images from the raw YFP and Caltech datasets trained by using synthetic image datasets. The investigation displayed the efficacy of the proposed

method for the detection of facial palsy with a significant increase ($p < 0.001$) compared to the baseline results. In addition, a better performance outcome was achieved by the proposed model in comparison to the existing models using the same YFP dataset. Further research would be beneficial to investigate other cutting-edge data augmentation techniques to create augmented datasets and create a reliable classifier with little computing complexity by merging models of transfer learning for rapid recognition of the moderate severity level of face palsy. In addition, by employing several face palsy datasets and cross-dataset validation of the proposed method, the robustness of the proposed approach shall be further investigated.

5.2. Analysis of Cassava Disease Detection

This study was implemented on a workstation running 64-bit *Windows 10* with an *Intel Core i5-8265U* CPU running at 1.60 GHz and 8GB of RAM. The suggested model was implemented by using the *MATLAB* (*MathWorks Inc*., USA) *Deep Learning Toolbox TM*. The network was trained for 10 epochs by using the initial learning rate of 0.001, a piecewise learning rate schedule, a learning rate drop factor of 0.2, and a learning rate drop period of 2 on the cassava leaf disease dataset using the Adam optimizer as summarized in Tbl. 5.6.

Table 5. 6. Model parameters for cassava disease detection

Model Parameters	
Batch size	100
Optimizer	Adam
MaxEpochs	10
WeightLearnRateFactor	10
LearnRateFactor	0.2
InitialLearnRate	0.001

5.2.1. Baseline implementation using transfer learning

The dataset was divided into 80% for training and 20% for validation. The trained model was tested on the validation set and received results of 0.977 accuracy, 0.9676 F1-Score, 0.9772 precision, and 0.9634 recall. The CMD class achieved the highest classification accuracy (98.5%), whereas the healthy class only received an accuracy of as little as 95.2%.

T-distributed *Stochastic Neighbor Embedding* (t-SNE) and network activations were used to analyze the output of a neural network. Fig. 5.4 (a) displays the FC layer's learnt features. The visualization of the network layers can enhance the interpretation of the classification process, and the outcome presents the learned features by the network using different gradients, as well as colors like yellow, green, red and orange spots. Fig. 5.4 (b) shows the network activations of the FC layer with the use of t-SNE to reduce dimensionality. There is a clear difference in each cluster representing the classes of the investigated plant disease. Thus, for this study, the trained network has been chosen as a baseline network model.

88
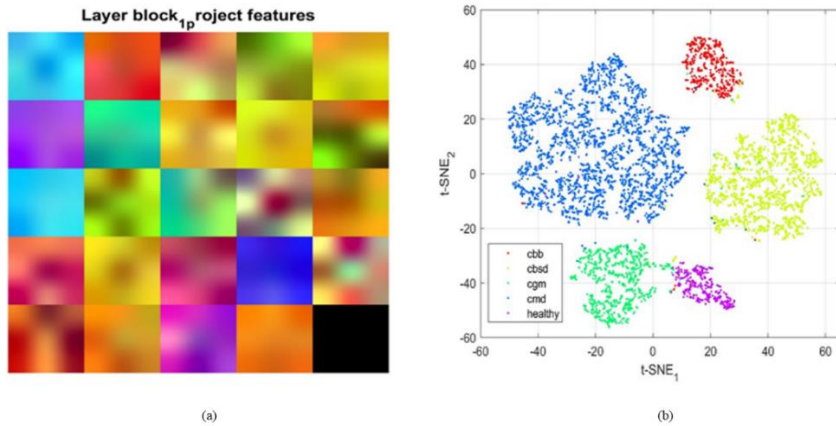
Fig. 5. 4. (a) The baseline neural network features images (15 channels and empty 16-channel); (b) use of t-SNE to activate the FC layer

5.2.2. Experimental results based on lower-quality images

The experimental test outcome using lower-quality dataset images is depicted in Fig. 5.5 (a)–(d). It can be deduced that, as the quality of the image deteriorates, the accuracy progressively declines. Fig. 5.5(a) shows that the classification accuracy rises as the image resolution increases, with values of 8%, 43%, 74.4%, 83.6%, 86.1%, and 87.3% for 32×32, 64×64, 96×96, 128×128, 160×160, and 192×192, respectively.
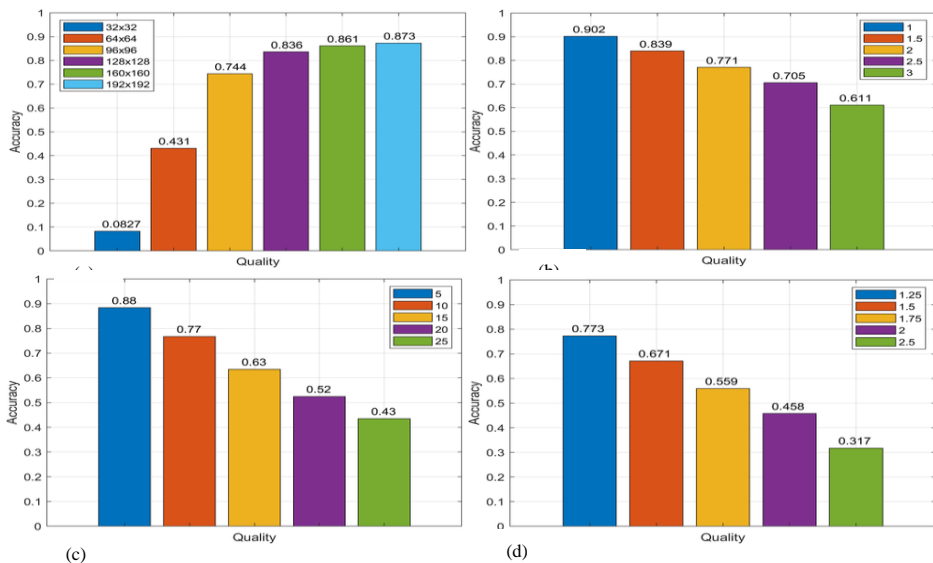
Fig. 5. 5. Performance outcome when using lower quality images: (a) Resolution-down-sampling; (b) Gaussian blurring; (c) Motion blur; (d) Overexposure

Furthermore, Fig. 5.5(b) presents images with Gaussian blur, and the classification accuracy declines slowly from 90.2%, 82.9%, 77.1%, 70.5%, and

61.1%, respectively, as the standard deviation parameter of Gaussian blurring σ is increased to the values of (1, 1.5, 2, 2.5, 3). In addition, Fig. 5.5(c) demonstrates that as the motion-blurring parameter was increased to 5, 10, 15, 20, and 25 pixels, the overall classification accuracy for motion-blurred images decreased, and the accuracy declined from 88%, 77%, 63%, 52%, to 43%, respectively. Fig. 5.5(d), where the accuracy rate is 77.3%, 67.1%, 55.9%, 45.8%, and 31.7%, respectively, shows how the improvement in overexposure varies with a persistent drop in the performance accuracy. In conclusion, our findings demonstrate that a decline in the quality of test images has an undesirable effect on the neural network's overall performance, thus indicating that the neural network struggles to generalize over images with resolution-down-sampling, Gaussian-blurring, motion-blur, and overexposure. This inspired the addition of the low-quality artificially created plant images to the raw dataset to train the network to identify leaf diseases in low-quality images as well.

5.2.3 Experimental results using augmented training dataset

In this instance, using the combination of both the raw training dataset and synthetic histogram-transformed images, the neural network model was efficiently trained. The combination of both the enhanced image dataset and the raw images contains 94,350 images and it was utilized to retrain the baseline network, with 80% of the images applied for training and 20% used for validation.
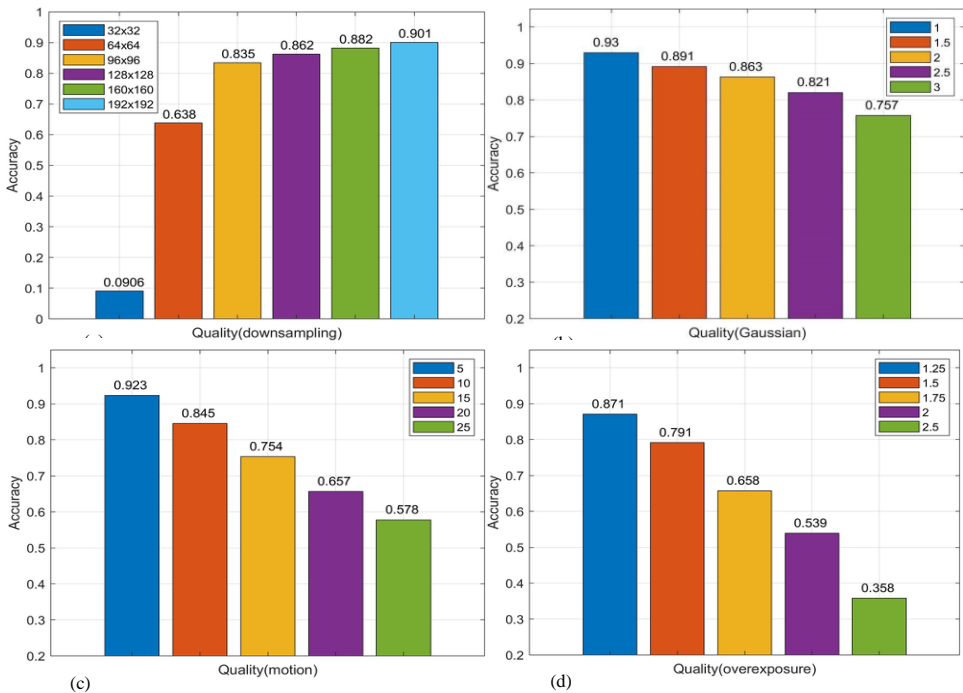
Fig. 5. 6. Performance outcome for augmented training dataset @ k= 10: (a) Resolution-down-sampling; (b) Gaussian blurring; (c) Motion blur; (d) Overexposure

The retrained network model was tested on the raw data, high-quality, and lower-quality cassava images. On the higher-quality images, the overall performance was able to attain an accuracy of 0.997, which is an improvement of 2%, however, the improvement was more pronounced regarding the lower-quality images. When down-sampling the resolution, the accuracy increased from 2.8% (192 × 192 px down-sampling) to 20.6% (64 × 64 px down-sampling). The accuracy of Gaussian blurring increased from 2.8% (blurred at 1σ) to 14.6% (blurred at 3σ). From 4% (motion blur at 5 px) to 14.3% (motion blur at 25 px), the accuracy of motion blurring increased. The accuracy for overexposure increased from 4% (2.5 times overexposed) to 9.8% (25 times overexposed). Fig. 5.6 (a) to (d), respectively, demonstrates the results of classification for a network evaluated on images with applied Gaussian blurring, motion blurring, and overexposure.

The network trained on an expanded image dataset generalized better on degraded synthetic images as compared to the outcomes of the baseline classifier trained on the original dataset images thus, the proposed model demonstrated better generalization abilities.

5.2.4. Statistical evaluation and outcomes overview

A nonparametric statistical Friedman's test and K-fold cross-validation were both used to assess the statistical significance with k = 10. When the ranking is random, the Chi-square distribution usually best describes the distribution of the data columns' mean ranks, which are what is determined for the test. To determine if the mean of the difference is equal to zero, the one-sample T-test was performed to the compute the difference in the classification accuracy. Tbl. 5.8 presents the results obtained.

The method's drawbacks are also indicated in Tbl. 5.8. The procedure did not provide any benefit when the image quality deterioration was considerable, as in the case of resolution down-sampling to $32 × 32$ pxg. The gain in the classification accuracy was more subtle for small image quality reductions (such as resolution-down-sampling to $192 × 192$ px, which is only marginally lower than the acceptable input size for the neural network). The outcome of the statistical analysis of the performance of the classifier shows that, in the majority of cases, the proposed data augmentation considerably increases the neural network's ability to identify cassava leaf disease.

Table 5. 7. Test accuracy on cassava disease dataset with statistical significance for each quality reduction parameter. Bold fonts convey the best values

Quality reduction method	Quality reduction parameters	Accuracy (Baseline)	Accuracy (Augmented dataset)	Improvement of accuracy	Results of Friedman's test	Results of one-sample t-test
Resolution down-sampling	32 × 32	0.0827±0.0146	0.0906±0.0164	0.0079±0.0158	NS	NS
	64 × 64	0.43147±0.0237	0.6378±0.0283	0.2064±0.0191	***	***
	96 × 96	0.7438±0.0291	0.8347±0.0203	0.0909±0.0249	***	***
	128 × 128	0.0836±0.0147	0.8821±0.0164	0.0262±0.0244	*	**
	160 × 160	0.0861±0.0201	0.8821±0.144	0.0209±0.0244	NS	*
	192 × 192	0.0875±0.0198	0.9006±0.0085	0.0281±0.0232	**	**
Gaussian blurring	1	**0.9017±0.0119**	**0.9298±0.085**	0.0281±0.0137	***	**
	1.5	0.8394±0.0175	0.8914±0.0147	0.0520±0.0263	***	***
	2	0.7705±0.0240	0.8628±0.0197	0.0922±0.0206	***	***
	2.5	0.7054±0.0218	0.8206±0.0169	0.1153±0.0232	***	***
	3	0.6113±0.0153	0.7573±0.0228	0.1460±0.0245	***	***
Motion blur	5	0.8834±0.0168	0.9234±0.0109	0.040±0.017	***	***
	10	0.7671±0.0099	0.8453±0.0172	0.078±0.018	***	***
	15	0.6346±0.0356	0.7536±0.0199	0.1190±0.0374	***	***
	20	0.5244±0.0258	0.6566±0.0199	0.1322±0.0217	***	***
	25	0.4346±0.0230	0.5776±0.0273	0.1431±0.0261	***	***
Overexposure	1.25	0.7727±0.0191	0.8710±0.0256	0.0983±0.0266	***	***
	1.5	0.6709±0.0255	0.9006±0.0287	0.1206±0.0243	***	***
	1.75	0.5588±0.0256	0.6577±0.0164	0.0262±0.0244	***	***
	2	0.4576±0.0228	0.5390±0.0302	0.0814±0.0220	***	***
	1.5	0.3175±0.0189	0.3580±0.0278	0.040±0.027	**	**

5.2.5. Evaluation and discussion

To verify the viability of the proposed strategy, an assessment of our proposed methods in comparison to that of [108] was conducted. The overall findings of this study were compared with the three best results from previous literature. Tbl. 5.7 illustrates existing methods, augmentation techniques, and experimental outcomes presented in the selected literature.

Table 5. 8. Comparison of performance results with existing methods for cassava disease detection

Ref.	Methods	Data augmentation	DNN Architecture	Accuracy
[108]	Proposed by Appian	Random crop, random erasing, horizontal and vertical flip, random affine	Se_ResNet	0.935
	Proposed by M. Vafan	Standard method	Se_ResNet	0.936
	Proposed by DeepBlueAI	Random crop, Horizontal flip, Vertical flip, rotate	Se_ResNet	0.939
[165]	Transformer (ViT) model based on self-attention and LeIAP	Flipping, zooming, cropping	ViT	-
Our Proposed Method		No augmentation	MobileNetV2	0.977
		Image color histogram equalization and IQRM	MobileNetV2	0.997

For evaluation, a comparison of our proposed work with related studies was further validated. Our approach was tested by using both high-quality images and synthetic low-quality images. The results are shown in Tbl. 23, and they demonstrate that our model provided the best performance, by improving the accuracy up to 0.977 when using it without any data augmentation and to 0.997 when using it in conjunction with the proposed approach for data augmentation. The impact of underfitting or overfitting may have had an influence on the results, particularly in the case of limited data for training; see [260]. To avoid overfitting, the MobileNetV2 neural network features built-in dropout layers. The effectiveness of such dropout layers has been demonstrated in [261]. In addition, the proposed image augmentation technique enables a significant increase in the number of images that are accessible for training, thus reducing the risk of underfitting. The proposed color-based transformation approach incorporated in creating extra training images can be viewed as a type of a regularization strategy which introduces variation to the original data to boost the training effectiveness [36]. Additionally, the *MobileNetV2* network used in this study features built-in batch normalization layers that are successful in preventing overfitting [260]. Although the model's complexity can lead to overfitting, it should be noted that using the first 18 layers of the model and adding three new layers greatly decreases the number of parameters in the MobileNetV2 complexity.

5.2.6. Conclusion of the section

The influence of data augmentation methods on deep neural networks and the experimental analysis of the created images with the image color space was presented. Convolution of the probability density functions of color histograms with Chebyshev orthogonal functions was presented as a further modification. To artificially reduce the image quality, four techniques were employed: resolution down-sampling, Gaussian blurring, motion blur, and overexposure. The *MobileNetV2* deep neural network was used for the classification. On high-quality images from the cassava plant disease dataset, the baseline network attained an accuracy of 97.7%; however, on low-quality images, the performance was noticeably lower, occasionally nearing that of a random classifier. The network was subsequently retrained by using both authentic (high-quality) and artificial images. When compared to a baseline neural network, this led to an increase in the classification accuracy for lower-quality photos from 3% to 15% while maintaining an accuracy of 99.7% for the original high-quality images.

The huge improvement in the classification rate can be attributed to the ability of the proposed augmentation techniques to generate realistic images with transformation emphasis on the color of the leaves. These augmented images were able to improve the ability of the generalization of learning of the model process and reduce overfitting. In spite of that, it is generally accepted that using high-quality images for classification would result in better prediction results. This study demonstrates the critical need to consider low-quality images to focus on the requirements of the intended users of AI-based applications, such as, for instance, rural farmers utilizing smart agriculture applications. This study's more general conclusion is the requirement for new benchmarks for evaluating and contrasting the effectiveness of NN models on corrupted test data.

Further research should concentrate on multi-class detection for locating various plant diseases. In addition, there is a need to enhance the time complexity of our proposed model's simplicity of implementation into mobile applications for real-time accessibility.

5.3. Analysis of Skin Melanoma Classification

An *Intel Corei5-5300U* CPU running at 2.30 GHz with 8GB RAM and a 64-bit version of *Windows 10* was used for this study. *MATLAB* (*MathWorks Inc.*, USA) *Deep Learning Toolbox* was implemented, and a *Stochastic Gradient Descent* (SGD) optimizer was used to train the network with over 10 epochs. The learning rate drop factor was set to 0.2, the learning rate drop period was set to 2, and the optimal learning rate was selected from a range of $10{-}4$, $10{-}5$, and $10{-}6$.

5.3.1. Classification results for binary-class

In the original dataset (PH2), the first test was run by using a 60:40 training-to-testing ratio. The original dataset is categorized into two sections which are the binary classification (melanoma and non-melanoma) and multi-class (melanoma, atypical, and common nevu classes). A total of 120 sample images were trained comprising 24 melanoma and 96 non-melanoma images (48 Common nevu and 48 Atypical nevu class). The test samples consisted of 80 test images, including 16 melanoma and 64

non-melanoma images (comprising 32 common nevu and 32 atypical nevu). The proposed CovSMOTE augmentation was used to produce AugDB-1 while AugDB-2 was created by using conventional transformation techniques, such as horizontal flipping, scaling, rotation, translation, etc. The results of the experiment revealed a significant result with an accuracy rate of 95.31% for AugDB-1, while there is almost a tie between the overall accuracy yield for the AugDB-2 and the original dataset with 89.26% and 89.64%, respectively. The application of our proposed model was able to reduce overfitting and improve the classification model's ability to generalize better. The experiment was carried out ten times, and the average performance metrics were 0.8964 for accuracy, 0.726 for sensitivity, 0.9406 for specificity, and 0.7365 for F1-score as demonstrated in Tbl. 5.9.

Table 5. 9. Best-performing results of the proposed method for Skin melanoma detection (Binary classification). Bold fonts convey the best values.

Binary Class Classification Results						
Dataset	**Description**	**Acc (%)**	**Pre (%)**	**Sen (%)**	**Spec (%)**	**F1-Score (%)**
Original	PH^2	89.64	76.12	72.58	94.06	73.65
AugDB-1	CovSMOTE Augmentation	**95.31**	**81.13**	**80.77**	**95.1**	**80.84**
AugDB-2	Conventional Augmentation	89.26	82.87	58.75	96.88	67.83

The pre-trained network was used to train and to build our proposed model, the proposed approach performed with the highest accuracy of the binary classification of 0.9531, a sensitivity of 0.8113, a specificity of 0.951, and an F1-score of 0.8084, as depicted explicitly in Fig. 5.7, while Tbl. 5.10 shows the confusion matrix summary of the test results for each training model built on the datasets which are the original image with no augmentation, covsmote, and conventional augmentation.

Table 5. 10. Confusion matrix with the influence of augmentation on the test dataset (Binary classification).

Original Image (No_Augmentation)				
Predicted/ Actual	Melanoma (Predicted)	Non-Melanoma (Predicted)	True Positive (%)	False Positive (%)
Melanoma (Actual)	10	3	62.5	37.5
Non-Melanoma (Actual)	6	61	95.3.0	9.0
AugDB-1 (COVSMOTE)				
Predicted/ Actual	Melanoma (Predicted)	Non-Melanoma (Predicted)	True Positive (%)	False Positive (%)
Melanoma (Actual)	15	3	93.75	6.25
Non-Melanoma (Actual)	1	61	95.31	4.69
AugDB-2				
Predicted/ Actual	Melanoma (Predicted)	Non-Melanoma (Predicted)	True Positive (%)	False Positive (%)
Melanoma (Actual)	14	4	87.5	12.5
Non-Melanoma (Actual)	2	60	93.8	6.2
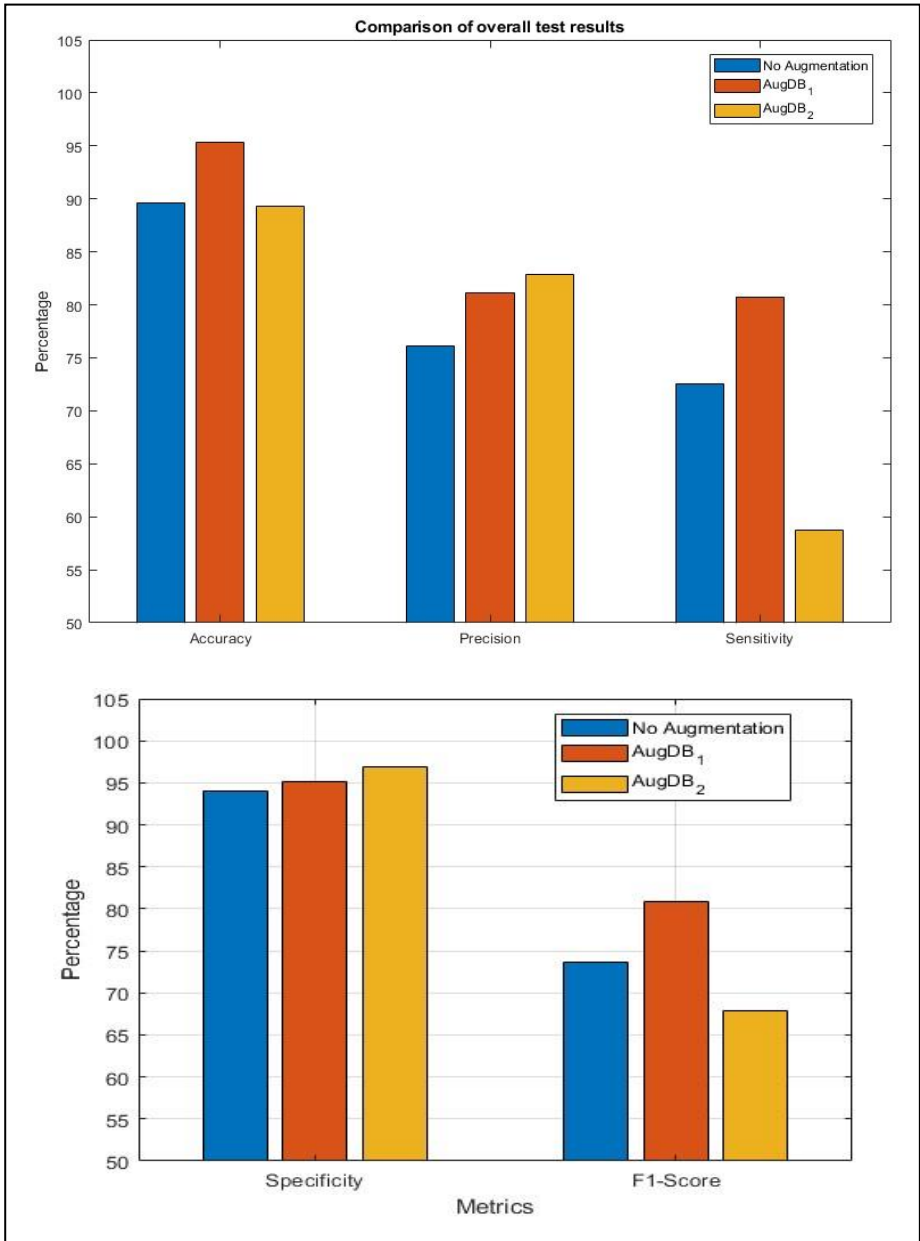
Fig. 5. 7. *Comparison performance for binary classification scenario*

5.3.2. Classification results for multi-class skin melanoma detection

A further experiment was performed for multi-class detection for the initial three classes of the PH2 dataset to obtain optimal results. By using the split ratio of 60 (training): 40 (test), new synthetic images were produced to balance the three classes: melanoma, common nevu, and atypical nevus, respectively. The outcome of multiple trials demonstrated that the mean accuracy rate, the sensitivity rate, the specificity rate, and the F1-score were computed. The test results, as presented in Tbl. 5.11, showed that, for the AugDB-1 dataset, the performance significantly enhanced the classification ability for melanoma, with the best accuracy of 0.694, a sensitivity of 0.892, a specificity of 0.962, a precision of 0.841, and an F1-score of 0,749, respectively. Fig. 5.8 presents a summary of the experimental findings for the three classes of melanoma, atypical nevu, and common nevu, with the performance comparison findings for the three databases. Tbl. 5.12 shows the confusion matrix for the outcomes of our proposed classifier.
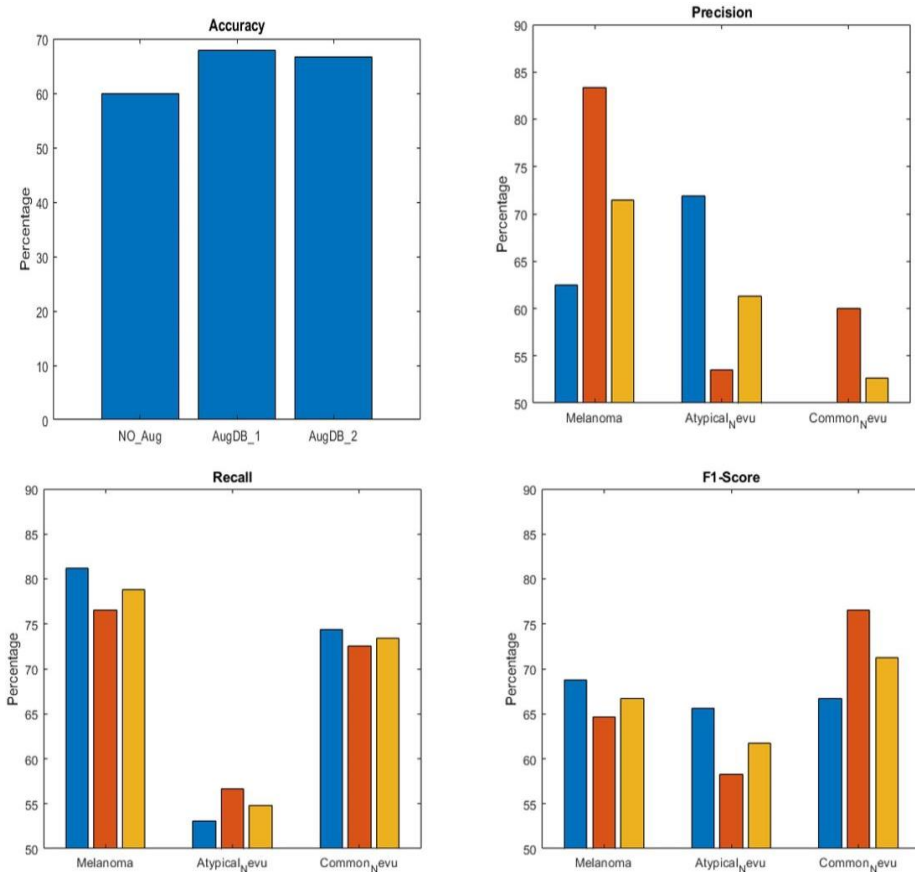
Fig. 5. 8. Comparison results for the three datasets in multiclass classification showing the accuracy (%), precision (%), recall (%), and F1-score (%)

Table 5. 11. Best-performing result for skin melanoma detection (multi-class)

Dataset	Accuracy	Melanoma				Atypical Nevu				Common Nevu			
		Sensitivity	Specificity	Precision	F1-Score	Sensitivity	Specificity	Precision	F1-Score	Sensitivity	Specificity	Precision	F1-Score
Original DB	60.4	64.2	93.6	73.8	67.4	54.8	68.9	68.9	53.8	63.9	73.4	62.0	62.5
AugDB-1	69.4	89.2	96.2	84.1	74.9	65.4	72.2	61.3	62.5	66.0	77.2	66.6	65.4
AugDB-2	66.0	61.5	73.0	76.1	57.6	57.4	61.9	54.0	58.4	56.1	75.2	64.0	58.3

Table 5. 12. Confusion matrix with the influence of augmentation on test dataset (multi-class)

Original Image (No_Aug)					
Predicted/ Actual	Atypical Nevu (Predicted)	Common Nevu (Predicted)	Melanoma (Predicted)	True Rate (%)	False rate (%)
Atypical Nevu (Actual)	27	3	2	84.4	15.6
Common Nevu (Actual)	19	12	1	35.5	62.5
Melanoma (Actual)	4	0	12	75.0	25.0
AugDB-1 (COVSMOTE)					
Predicted/ Actual	Atypical Nevu (Predicted)	Common Nevu (Predicted)	Melanoma (Predicted)	True Rate (%)	FP (%)
Atypical Nevu (Actual)	19	11	2	59.38	40.63
Common Nevu (Actual)	10	21	1	65.63	34.38
Melanoma (Actual)	3	0	13	81.25	18.75
AugDB-2					
Predicted/ Actual	Atypical Nevu (Predicted)	Common Nevu (Predicted)	Melanoma (Predicted)	True Rate (%)	FP (%)
Atypical Nevu (Actual)	20	9	3	62.5	37.5
Common Nevu (Actual)	11	21	0	65.6	34.4
Melanoma (Actual)	4	1	11	68.75	31.25

5.3.3. Benchmarking with previous work

A comparison of the performance of the proposed approach by using the PH2 dataset and the state-of-the-art methods from an earlier-published work is shown in Tbl. 5.13. In a few instances, our proposed approach significantly outperforms certain prior research. The proposed approach had relatively small computational requirements given the temporal complexity of both binary and multiclass problems. In comparison to some of the findings from the earlier investigations, our proposed approach performed better. The proposed method's shortcoming is that the sensitivity and specificity explanation rate is relatively low and requires further improvement via investigating different deep learning architectures in the future.

Table 5. 13. Comparison of performance results with existing methods by using PH2 dataset. Bold fonts convey the best values

Method	Acc (%)	Sen (%)	Spec (%)	Reference
Combined Gaussian method and color space transformation. Inception-V3 classifier	97.74	97.39	98.10	[222]
Feature Similarity Measurement (FSM) algorithm on SVM-classifier	91.90	92.50	91.30	[262]
SVM-classifier with the linear kernel	86.07	78.93	93.25	[263]
Markov Random Field and Fuzzy C-Means methods	94.0	93.20	98.0	[264]
Joint Reverse Classification and Multi-scale Lesion biased Rep	92.0	87.50	93.13	[180]
Deep CNN	**98.50**	93.0	100	[265]
SqueezeNet and CovSMOTE Augmentation	92.18	80.77	95.10	**Proposed Method**

5.3.4. Conclusion of the section

The majority of the cases of skin cancer mortality is often a result of melanoma, and the present research focuses on an early diagnosis of this condition. The expansion of the research efforts in this field has been hampered by several obstacles, including a small amount of data for skin cancer detection, an imbalance in the classes of skin tumors, a lack of labeled data, poor clinical image standardization, etc., which has led to the poor performance of the classifier. By carefully implementing the proposed approach, this study enhanced the classification performance of melanoma skin disease based on these difficulties. The proposed approach is a successful data augmentation technique based on covariant SMOTE to address the issue of class imbalance. Through comparison with other currently existing methods and the conventional DAM techniques, the usefulness of the proposed data augmentation method has been demonstrated.

The outcomes of multiclass classification scenarios show that the designed NN architecture could assist medical professionals screen patients for melanoma

effectively. Through comparisons with other methods, the proposed method's efficacy and dependability have been demonstrated. The proposed strategy has shown promising results, it can be successfully included in a device with limited resources, and it effectively addresses the issue of real-time skin melanoma detection.

The future study direction is to investigate potential paths by combining various DL architectures, such as AlexNet, ResNet101, and DenseNet201, to enhance our proposed data augmentation technique for the effective identification of melanoma disease. To alleviate the current research bottleneck in small data analytics, the proposed image enhancement technique can also be used in other imaging domains, such as plant disease diagnosis, defect detection, etc. Finally, the intention of presenting our model in a smartphone app seems viable so that to create an AI-based decision support system which would assist users in performing self-clinical examinations, which will help with the early diagnosis of skin cancer disorders.

5.4. Analysis of COVID-19 Detection (Sound Dataset)

5.4.1. Performance results and discussion

A detailed analysis of each dataset on the proposed *DeepShufNet* was considered along with the condition of the hardware requirements. Multiple experiments were conducted on all the different datasets on the proposed DeepShufNet. All experiments were conducted in *MATLAB R2020b* on a desktop PC built with an *Intel(R)* core *i5* (3.2 GHz) processor, 8 GB of RAM, and an *NVIDIA GeForce GTX 1070* GPU server with 120 G memory. An adaptive momentum algorithm *ADAM* method was used to adaptively alter the learning rate to a value of 0.0001. For the optimal results, a batch size of 200 was used, a dropout of 50% was adopted, and the maximum epoch to be used was 50. The summarized parameters are listed in Tbl. 5.14.

Table 5. 14. Configuration Requirements and Parameters for Sound Classification

Model Parameters	
Batch size	200
Optimizer	Adam
'MaxEpochs'	50
'WeightLearnRateFactor'	10
'BiasLearnRateFactor'	10
'InitialLearnRate'	0.0001

5.4.2. Classification of the proposed model: DeepShufNet performance

The combined feature-extracting images from all of the *COSWARA* datasets were used to train and test the proposed *DeepShufNet* model. The model was trained and validated by using 80% of the data, and the remaining 20% was used for testing the model. The training optimization was based on an adaptive momentum algorithm, with the hyperparameter values of the batch size of 200, a 50% dropout, and the epoch level 50. While the batch size aids in determining the sample, the rows processed, and the duration before changing the network's internal parameters, the learning rate

regulates the rate of weight updates, hence minimizing the prediction error. By using images from the raw feature extraction from the baseline experiment, the training process was assessed both with/without fine-tuning. The DeepShuffleNet model with the lowest loss in the validation set during training was chosen as the final DeepShuffleNet model.

For every experiment, analysis and observations were implemented for the overall outcome of the DeepShufNet model using the validation accuracy and losses. The original dataset's results struggle with a rising minority class misclassification rate, particularly when positive asymptotic and positive COVID-19 classes are classified with recall and precision rates of nearly NA to less than 10%. The DeepShufNet model performed better at detecting COVID-19 after being trained with our categories of the synthetic dataset. All of the results were derived based on trials with the test dataset and are provided in four comparison categories. In five recorded trials, the model's overall effectiveness for every category of datasets is compared by using an ideal model. Since the outcomes of both classes are examined in each comparative experiment, and the minority class classification outcomes are improved, the accuracy, recall, and the specificity combination serve as the primary metric to assess the model's performance in each dataset.

5.4.3. Experimental result 1: (all positive COVID vs. healthy)

The positive-mild and the positive-moderate classes were combined due to the similarities or common feature attributes between both classes and were thus used to establish a new class named the 'All-positive-covid class', considering their commonalities. The experimental outcome of the *DeepShufNet* model on 224×224 pixels for binary classification was presented, and Tbl. 5.15 displays the outcomes of our proposed DeepShufNet potential to identify Mel spectrogram feature images and GFCC features. The classification outcomes demonstrate a balance and efficiency of DeepShuffleNet in the data augmentation datasets.

Table 5. 15. Average-performing results for COVID-19 detection using the COSWARA dataset (all positive COVID-19 vs. healthy)

Exp 1: Positive COVID vs. healthy						
Features Extraction	Dataset	Acc (%)	Rec (%)	Prec (%)	Spec (%)	F1-Score (%)
Mel Spectrogram	No-Aug	71.2±7.3	60.2±12.6	51.8±15.1	85.22±11.3	53.47±6.1
	COCOA_1	78.7±6.1	57.9±13.5	45.41±9.6	83.19±9.4	49.2±5.8
	COCOA_2	85.1±4.2	70.85±7.7	59.64±13.1	88.25±6.14	63.61±6.7
	COCOA_3	87.8±1.3	69.49±4.9	64.82±4.7	91.75±1.9	66.9±2.8
GFCC	No-Aug	74.9±3.8	48.7±14.1	40.1±10.16	86.99±1.55	42.4±6.3
	COCOA_1	76.4±2.5	71.33±2.2	41.23±3.4	77.51±3.3	52.17±2.6
	COCOA_2	83.1±0.9	46.7±11.5	53.3±2.32	91.06±2.01	49.27±6.5
	COCOA_3	83.1±1.4	38.33±9.3	50.21±1.6	92.21±1.7	43.1±6.5

The Mel spectrogram images COCOA-2 datasets yielded the best results for DeepShufNet on the test set, with an exciting positive-COVID classification instance computed as the average accuracy with *Standard Deviation* (SD) as 85.1 (SD = 4.23), 70.85 (SD = 7.7) for recall/sensitivity, 59.64 (SD = 13.12) for precision, 88.25 (SD = 6.14) for specificity, and 63.61 (SD = 6.7) for F1-score. The results from the test set for our proposed model on COCOA-3 showed a considerable increase in accuracy, with an average of 87.82 (SD = 1.3), 69.49 (SD = 4.9) for recall/sensitivity, 64.82 (SD = 4.7) for precision, 91.75 (SD = 1.9) for specificity, and an F1 score of 66.9 (SD = 2.8). As a result, when compared to the results of the other datasets, it can be said that the test set results against the baseline results perform the least effectively, as presented in Fig. 5.9.

For the GFCC images, the augmented dataset achieved better results than the baseline results in terms of the accuracy with the scores of 83.1% (SD = 1.4) for COCOA-3, 83.05% (SD = 0.9) for COCOA-2, 76.4% (SD = 2.5) for COCOA-1, and 74.9% (SD = 3.8) for the original data (see Tbl. 5.15). The growing mean recall for DeepShufNet is more intriguing with a rate of 71.3% (SD = 2.2) for COCOA-1, 48.7% (SD = 14.1) for the original data, 46.7% (SD = 11.5) for COCOA-2, and 38.8% (SD = 9.3) for COCOA-3. Fig. 5.10 shows the comparison test results of the four databases and the improvement in the detection rate across all the metrics by training with augmented data.

The analysis of DeepShufNet on Mel spectrogram images reveals that COCOA-2 had the most promising experimental results, with scores of 90.1%, 62.71%, 95.99%, 77.1%, and 69.2% for the accuracy, recall, specificity, and precision, respectively. The next-to-best results were obtained by COCOA-3 with an 89.5% accuracy rate, a recall rate of 71.2%, a specificity rate of 93.4%, a precision rate of 70%, and an F1-Score rate of 70.6%. The original dataset without augmentation produced the least accurate results (79% accuracy, 54.23% recall, 84.3% specificity, 42.67% precision, and 47.76% F1-score). Similar results were obtained for GFCC images using COCOA-2 and COCOA-3, which, respectively, achieved an impressive accuracy of 84.1% and 84.7%. The implementation of a data augmentation strategy helped to enhance the performance outcomes, as seen by the two COCOA-1 and COCOA-2 results which had the highest recall rate.

5.4.4. Experimental results 2: (positive asymptotic vs. healthy)

To further validate the performance, a second experiment was carried out to employ the proposed *DeepShufNet* models for distinguishing between healthy and positive asymptotic alone. The performance metrics for both Melspectrogram and GFCC images are not continuous for the original dataset, even if the application of the DAM strategy on training data offered minimized overfitting with the training accuracy substantially reduced in comparison with the test results in the most recent epoch. Nevertheless, the training results with the model using COCOA-1 demonstrated better classification performance for the test sets in terms of accuracy.

As a result, according to the experimental results, not all training with enhanced datasets significantly improves the classification accuracy. In contrast to this, the DeepShufNet model with COCOA-2 improves the test results by increasing the accuracy, specificity, and F1-score rates. As shown in Tbl. 5.16, when the dataset is

very small, the classification performance utilizing noise augmentation is more acceptable when applied by practically taking the application efficiency of the DAMs into account. The experimental results demonstrate a significant improvement over the baseline experiment (no-Aug) when data augmentation is used, with COCOA-1 achieving the highest accuracy (97.15% SD, 0.5), followed by COCOA-2 at 95.8% SD, 1.1, COCOA-3 at 92.7% SD, 0.17, and No-Aug data at 92.2% SD, 0.9.

Table 5. 16. Average-performing results for COVID-19 detection using the COSWARA dataset (positive asymptotic vs. healthy)

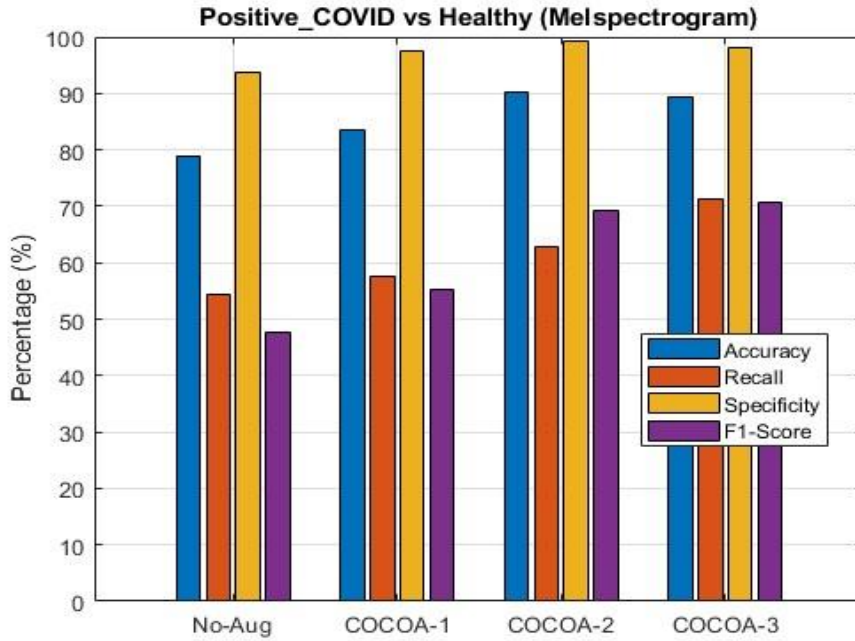| Exp 2: positive asymptotic vs. healthy						
Features Extraction	Dataset	Accuracy	Recall	Precision	Specificity	F1-Score
Mel Spectrogram	No-Aug	92.2±0.9	18.75±8.8	9.03±4.9	94.3±0.8	10.73±4.1
	COCOA-1	97.15±0.5	25±17.7	71.42±20.4	99.27±1.0	31.11±12.6
	COCOA-2	95.8±1.1	37.5±15.3	30.43±8.3	97.37±1.1	32.45±8.9
	COCOA-3	92.7±0.17	41.67±7.2	17.19±2.43	94.16±0.0	24.34±3.7
GFCC	No-Aug	88.54±0.9	12.5±3.3	12.5±1.2	93.79±0.77	12.18±6.35
	COCOA-1	91.15±5.4	NA	NA	93.79±5.4	NA
	COCOA-2	92.86±2.1	19.17±6.3	17.84±13.63	96.59±2.35	17.06±7.7
	COCOA-3	89.0±5.93	25.0±12.5	9.96±5.8	90.88±6.5	12.01±2.3

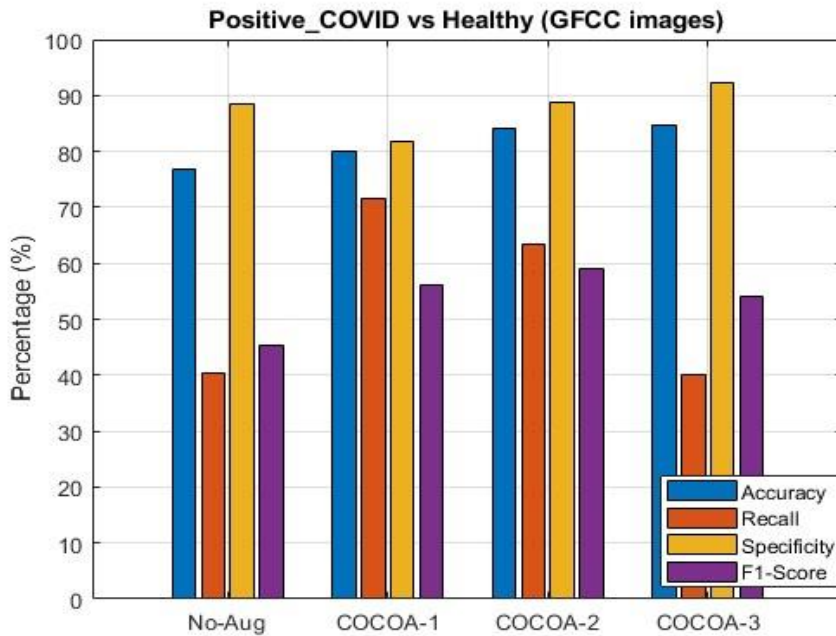**Fig. 5. 9.** Comparison results for Positive COVID vs. healthy Mel Spectrogram

Fig. 5. 10. Comparison results for Positive COVID vs. healthy (GFCC images)
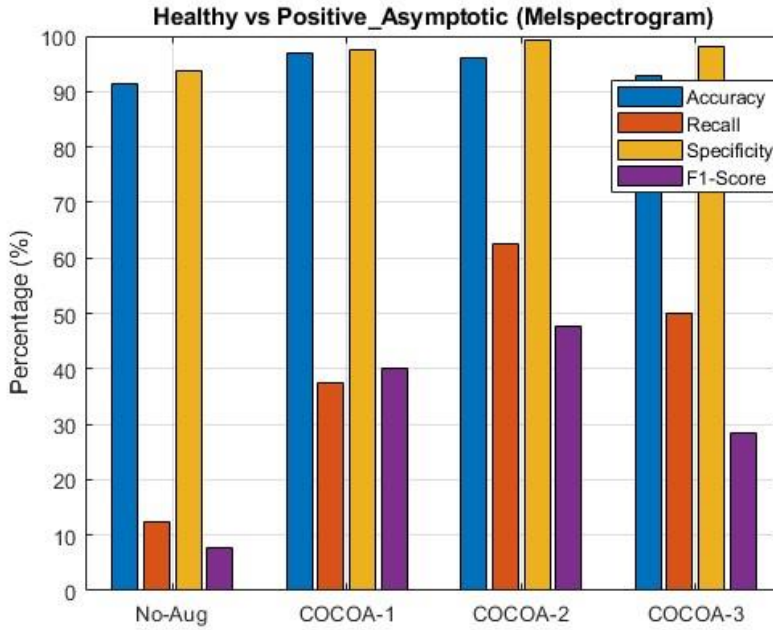
Fig. 5. 11. Comparison results for positive asymptotic vs. healthy (Mel Spectrogram)
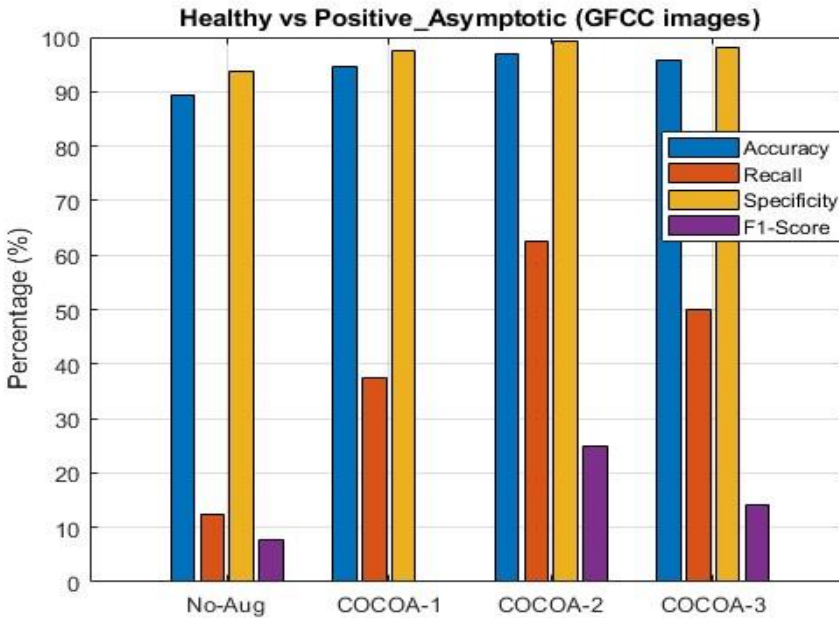
Fig. 5. 12. Comparison results for positive asymptotic vs. healthy (GFCC images)

106

The best metrics analysis of the different datasets and the results depict a clear recall rate, precision, and F1-score increase when Mel-spectrogram images are enhanced with more successful algorithms, as represented in Fig. 5.11 and Fig. 5.12. Therefore, it could be agreed that the proposed DeepShufNet model's classification results have improved significantly because of the data augmentation techniques used in both feature extraction images.

5.4.5. Experimental results 3: (healthy vs. recovered full)

Further investigation was conducted to examine the efficacy of the proposed *DeepShufNet* model in identifying the healthy-versus-recovered data. Fig. 5.13 and 5.14 show the comparison results between all the datasets and illustrate the best performance of the Mel-spectrogram. In addition, Tbl. 5.17 serves as an indication of the progress in the analysis of the models based on four datasets for the MFCC and GFCC feature images. The performance analysis was done for the original data (no augmentation), COCOA-1, COCOA-2, and COCOA-3, and the produced results showed an accuracy of 93.45 (SD, 0.41) for COCOA-2, 93.33 (SD, 0.51) for COCOA-1, 91.68 (SD, 4.0) for COCOA-3, and 91.03 (SD, 0.8). It also demonstrated that the best outcomes were obtained when using the combined DAM strategy which is referred to as COCOA-3.

Table 5. 17. Average-performing results for COVID-19 detection when using the COSWARA dataset (healthy vs. recovered full)

Exp 3: healthy vs. recovered full						
Features Extraction	Dataset	Accuracy	Recall	Precision	Specificity	F1-Score
Mel Spectrogram	No-Aug	91.03±0.80	96.96±2.70	94.56±0.60	26.32±0.00	95.72±0.9
	COCOA-1	93.33±0.51	98.78±1.10	94.31±0.40	18.33±7.60	96.49±0.3
	COCOA-2	93.45±0.41	96.42±1.20	96.5±1.00	52.5±14.40	96.45±0.2
	COCOA-3	91.68±4.00	92.89±5.60	97.14±1.40	61.25±20.60	95.39±2.4
GFCC	No-Aug	90.5±1.90	95.86±2.80	93.16±0.60	18.33±11.54	94.98±1.1
	COCOA-1	77.96±12.90	81.47±14.66	93.90±0.50	27.0±16.05	86.66±8.5
	COCOA-2	91.15±2.80	97.54±3.10	93.28±0.22	5.0±0.00	95.34±1.6
	COCOA-3	91.04±4.00	97.08±4.30	93.53±0.32	8.0±4.50	95.23±2.2

Fig. 5. 13. Comparison results for healthy vs. recovered full (Mel Spectrogram)

Fig. 5. 14. Comparison results for healthy vs. recovered full (GFCC images)

5.4.6. Comparison to alternative classification models

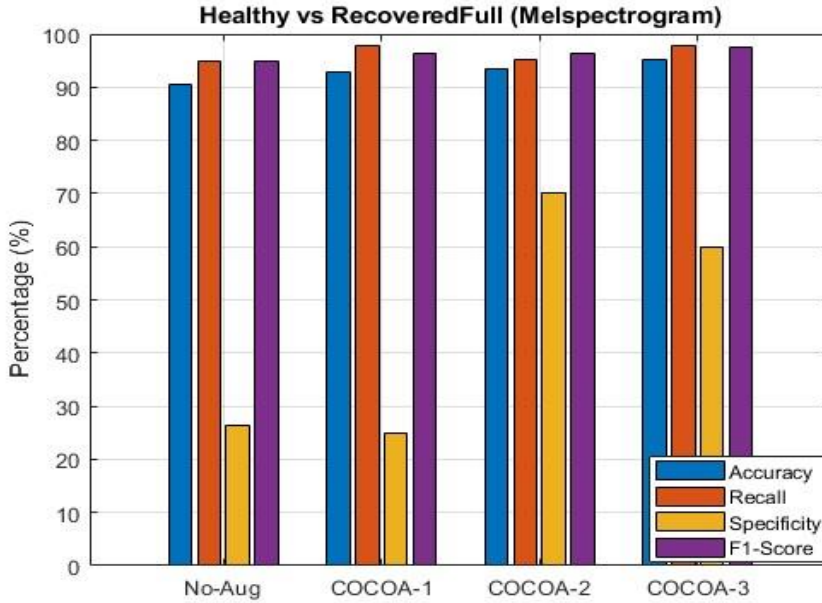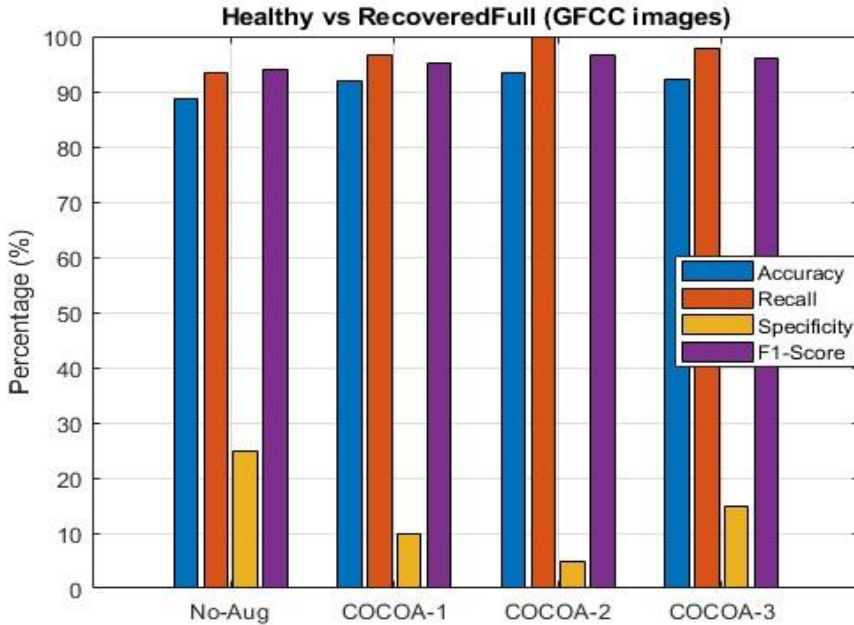Analyses were evaluated to compare our proposed deep learning results to other related approaches which have been used in previous studies using the COSWARA data to classify COVID-19. The proposed *DeepShufNet* model exhibits improved and promising outcomes for COVID-19 detection compared to the earlier studies while applying varied experimental circumstances to each classification task. Tbl. 5.18 clearly shows the summarized results in contrast to the currently existing techniques.

Table 5. 18. Comparison of performance results with the existing methods using COSWARA dataset

References	Classifier	Acc	Spec	Sens	AUC
[267]	Deep model shallow classifier	92	98	93	97.6
[268]	Ensemble DLmodel	77.1	NA	NA	NA
[269]	VGGish shallow	72	NA	NA	NA
[242]	ResNet50, SVM, LR, LSTM, KNN	91	92	90	92.3
Our work	DeepShufNet	**<u>90.1</u>**	95.98	62.3	NA

5.4.7. Drawbacks of the proposed model

The issue of misclassification mistakes linked to the poor generalization of some noisy images is among the study's most pressing problems. As anticipated, a significant imbalance of classes and few data samples can account for most of the misclassification errors. The capacity of the model to effectively generalize the data could be impacted by the similarities between the disparities between each class of sound and power representation when displayed as images, such as Mel-spectrogram or GFCC images. According to to Esmaeilpour et al. [266], the resulting spectrogram for each audio file is a 2D array of intensity values that is significantly noisy due to background sounds associated with audio signals. Therefore, it is crucial to improve the intensity values distributions to improve feature learning. The proposed framework is created by using prior data augmentation methods (such as color transformation and noise) and characteristics in the frequency domain, thus making it easy to understand and ensuring low space costs. Although previous studies have used the frequency-domain feature in sound classification tasks, given that some images cannot accurately represent the distinctive features of sound signals, image representations for sound signals could be a complicated system. Despite these drawbacks, the proposed DeepShufNet has demonstrated success in detecting COVID-19 despite the high imbalance classes and an insufficient dataset, and it is denoted by a lower computational complexity. To overcome some of the flaws caused by inaccurate classification of images and to create a more efficient dataset for proper generalization, more advanced data augmentation techniques will need to be investigated in the future.

5.4.8. Conclusion of the section

It is rather impressive how frequently various DNN models are used for sound classification tasks. Despite considerable research on COVID-19 identification involving the use of different CNN architectures, some publicly accessible datasets still suffer from data sparsity and class imbalance, which leads to poor classification from some ML models. This research intends to apply various kinds of data augmentation techniques to a deep learning model named *DeepShufNet*. Among this work's primary contributions are the following. By applying certain straightforward and efficient data augmentation strategies, this study was able to bridge the gap between the lack of sufficient datasets and the class imbalance. Three additional datasets, coined COCOA-1, COCOA-2, and COCOA-3 were created to train our proposed learning model. In addition, the proposed DeepShufNet model, which uses a pre-trained Shufflenet architecture, was trained and assessed by using the comparison datasets. A comparison of experimental analysis of the enriched datasets to the baseline findings revealed considerable improvements in the performance measures, better generalization, and improved optimal test results.

On the DeepShufNet model, the effects of the two alternative feature extraction techniques (GFCC image and Mel-spectrogram) were examined. In this study, the impact of enhanced photos on the detection of COVID-19, positive asymptotic cases, and fully recovered cases were examined. The outcomes demonstrated that, for nearly all the comparative situations, the DeepShufNet model had the best accuracy on COCOA-2 Mel-spectrogram images. For all the three types of enhanced images, the proposed DeepShufNet models displayed a significantly increased performance, particularly in the recall rate, precision, and the F1-Score rate. The proposed approach for detecting positive-COVID using the Mel COCOA-2 training datasets had the greatest test results for accuracy, precision, recall, specificity, and F1-score, respectively, of 90.1%, 77.1%, 62.7%, 95.98%, and 69.1%. Similar to that, the experimental findings for the detection of positive asymptotes had the best recall rate of 62.5%, a specificity rate of 97.1%, and an F1-score of 48%.

The application of the improved deep learning architectures could be used in future research to strengthen and improve the outcomes of the COVID-19 classification by using sound data. Also, the combination of all the various sound datasets available in the COSWARA database might be used to apply and analyze the proposed DeepShufNet model.

5.5. Analysis of all Proposed Augmentation Methods

The evaluation of the four proposed data augmentation methods in this dissertation, namely, VDRRE, CovSMOTE, Photometric (IQRM), and Noise Injection methods (NIM) has been analyzed and investigated on the pre-trained network. To provide a thorough assessment, a comparative evaluation was conducted contrasting the performance of the three effective pre-trained networks which include SqueezeNet, EfficientNet, and ResNet18, for identifying the unique features and then used to accurately classify facial palsy and skin melanoma diseases on the proposed augmentation methods. This comparison was conducted across two datasets so that to fully understand the uniqueness of each augmentation approach concerning their

strengths and shortcomings. Through this comprehensive investigation, more light has been shed on the peculiarity of the proposed DAM methods and their impact on the overall performance in developing better generalizations of deep learning models.

The hyperparameter settings used to implement and optimize the pre-trained SqueezeNet, EfficientNet and ResNet18 architectures were designed to suit the features of both datasets (PH2 and YFP). The entire setup for this experiment has been carefully fine-tuned to ensure that the model is trained to identify and categorize effectively the skin melanoma and facial palsy data diseases. The specification and its values are displayed in Tbl. 5.19.

Table 5. 19. Hyperparameter settings and their values

Parameters	Values		
	SqueezeNet	EfficientNet	ResNet18
Initial learning rate	0.00001	0.0001	0.003
Activation Function	ReLU	ReLU	ReLU
Epochs	50	3	10
Batch Size	16	100	100
Optimizer	SGDM	ADAM	ADAM
L2Regularization	0.01	0.1	0.01
Momentum	0.9	0.9	0.9
WeightLearnRateFactor	20	10	10
Trainable parameters (millions)	1.2	5.3	11.6
No. of layers	63	290	71

For skin melanoma disease detection, the test result for covSMOTE augmentation achieved the best accuracy results for the three classification models. Regarding Tbl. 5.20, the covSMOTE augmentation, based on ResNet18, achieved the best classification results in comparison with other augmentation test results in this category with an accuracy of 0.938, while VDRRE, PM, and NIM obtained an accuracy of 0.925, 0.90, and 0.912, respectively, on the test dataset. The pre-trained network SqueezeNet also showed significant results with the best performance obtained with the covSMOTE with an accuracy of 0.9218; NIM achieved an accuracy of 0.8522, PM achieved an accuracy of 0.8436, and the VDRRE accuracy of 0.775 was reached for the test dataset. The proposed covSMOTE augmentation method demonstrated impressive performance in accurately classifying skin melanoma disease by indicating that there is no bias against any sample with the overall recall, precision, specificity, and F1-score of 0.813, 0.867, 0.968, and 0.839, respectively. Overall, the proposed covSMOTE augmentation model was able to correctly classify more instances of the palsy positive class, as evidenced by the balance in accuracy, recall and precision. Therefore, it can be agreed that the training of the three models, especially (ResNet18 and SqeezeNet) with covSMOTE showed a good resilience and generalization ability in our model to identify and classify melanoma accurately.

Furthermore, in face palsy detection, the augmentation based on VDRRE in comparison with other proposed augmentation methods shows a significant variability across the three pre-trained networks. The experimental results summarized in Tbl. 37 clearly demonstrate the testing performance of the proposed DAM methods in the detection of face palsy. It can be deduced from Tbl. 5.21 that the test accuracy in the case of VDRRE techniques remains consistent with a smooth increase across the three pre-trained CNN architectures with the maximum accuracy reaching 0.9935 for squeezenet, while the test accuracy for the EfficientNet and ResNet architecture for the same VDRRE obtained the values of 0.965 and 0.929, respectively. However, the performance of the test data with covSMOTE augmentation repeatedly shows the lowest results in comparison with the other proposed DAM methods. However, five experimental runs, and fine-tuning by reducing the learning rate to 0.0003, aided the overall classification of the deep learning model, thus increasing the accuracy of the test result by +8.3%. In addition, the application of the photometric and NIM approach showed a somewhat conservative outcome, which was able to mitigate overfitting with the help of parameter regularization during the training of the model. In conclusion, the experimental results demonstrated very good capabilities to deal with small data problems through leveraging on the relevance of the domain-specific knowledge to successfully influence the model learning process and thus overcoming the shortcomings of small data in real-life applications. However, it can be argued that no single augmentation technique can be considered as the 'best' or 'perfect' method for all small data cases due to the uniqueness of the data features and the target domain.

Table 5. 20. Comparison of the proposed augmentation methods for skin melanoma classification

Skin Melanoma Classification						
Data Augmentation Methods (DAM)	Classification Model	Accuracy	Recall	Precision	Specificity	F1-Score
Voronoi Decomposition random region erasing (VDRRE)	SqueezeNet	0.775	0.6875	0.4479	0.7969	0.5249
Covariant SMOTE (covSMOTE)		0.921	0.8077	0.8257	0.951	0.8084
Photometric (Color Transformation)		0.8436	0.8021	0.6134	0.8542	0.6625
Noise Injection Method (NIM)		0.8522	0.6667	0.6588	0.8984	0.6352
Voronoi Decomposition random region erasing (VDRRE)	EfficientNet	0.862	0.50	0.7273	0.9531	0.5926
Covariant SMOTE (covSMOTE)		0.875	0.375	0.75	0.9688	0.500
Photometric (Color Transformation)		0.838	1.000	0.5517	0.7969	0.7111
Noise Injection Method (NIM)		0.85	0.625	0.625	0.9063	0.625
Voronoi Decomposition random region erasing (VDRRE)	ResNet18	0.925	0.875	0.968	0.667	0.919
Covariant SMOTE (covSMOTE)		0.938	0.813	0.867	0.968	0.839
Photometric (Color Transformation)		0.90	0.625	0.8333	0.9686	0.7142
Noise Injection Method (NIM)		0.912	0.737	0.875	0.9672	0.80

Table 5. 21. Comparison of the proposed augmentation methods for Face Palsy Disease Classification

Face Palsy Detection						
Data Augmentation Methods (DAM)	Classification Model	Accuracy	Recall	Precision	Specificity	F1-Score
Voronoi Decomposition random region erasing (VDRRE)	SqueezeNet	**0.9935**	**0.9974**	**0.9935**	**NA**	**0.9954**
Covariant SMOTE (covSMOTE)		0.662	0.525	1.00	1.00	0.688
Photometric (Color Transformation)		0.801	0.991	0.785	0.333	0.876
Noise Injection Method (NIM)		0.791	0.8959	0.825	0.5333	0.8590
Voronoi Decomposition random region erasing (VDRRE)	EfficientNet	0.965	0.9638	0.9861	0.9667	0.9748
Covariant SMOTE (covSMOTE)		0.675	0.7692	0.7727	0.4444	0.7710
Photometric (Color Transformation)		0.720	0.9321	0.7410	0.20	0.8826
Noise Injection Method (NIM)		0.875	0.8778	0.9417	0.8605	0.9087
Voronoi Decomposition random region erasing (VDRRE)	ResNet18	0.929	0.914	0.9854	0.9667	0.9483
Covariant SMOTE (covSMOTE)		0.579	0.407	1.000	1.000	0.5788
Photometric (Color Transformation)		0.624	0.4932	0.9561	0.9444	0.6507
Noise Injection Method (NIM)		0.768	0.7919	0.8706	0.7111	0.8294

5.6. Limitation of Methods

There are certain limitations that arise from this study due to the influence of the extremely small and imbalanced dataset, as well as dependency of deep learning models on big data for effective and better classification interactions. The covSMOTE approach showed some good results in the case of skin melanoma on different pre-trained models, however, it displayed relatively low results in the case of face palsy. As a matter of fact, this can be attributed to the overlapping of classes as the augmented images include noise, and thus the image may look unrealistic as a result. In addition, there was the issue of model overfitting in the skin melanoma detection scenario, as the test results showed some biases toward the minority class, especially for the binary classification task. Furthermore, in the multi-class scenario, although the model shows optimal results for accurately classifying melanoma from the other class, however, it shows misclassification error in the case of differentiating atypical nevu and common nevu classes, which results in a lower recall rate and a lower specificity rate. For the sound classification task, the main challenge to be faced was found to relate with detection errors which can be attributed to the poor generalization of some noisy images created. In addition, for each class of sound, it was realized that, after the feature extraction techniques have been applied, the featured images are closely similar in power representation, thus affecting the capability of the learning classifiers to effectively generalize data.

5.7. Experiments Summary

In this chapter, the findings and experimental results of applying different data augmentation models and learning classifiers in different classification tasks have been presented. By using the color histogram augmentation method, the learning model (MobileNetV2) was able to obtain an improved accuracy of 99.7% on high-quality images. Training with the Gaussian noise augmentation showed an overall best classification result. Further comparison of the performance with the currently existing methods shows that the proposed model is efficient for the detection of cassava disease. The outcome shows a statistically substantial improvement in the capability of cassava leaf disease detection shown by the enhanced MobileNetV2 neural network. A key aspect of practical data collecting is the model's ease of deployment for identifying and diagnosing cassava leaf diseases in images of lesser quality.

For facial palsy detection, implementation using the proposed covSMOTE model generated new synthetic datasets, and the overall performance on the test dataset has improved. The implemented shuffleNet and SVM detection algorithm has demonstrated better performance, with an improved accuracy, sensitivity/recall rate, specificity, and a lower time complexity. Finally, our proposed model provided better insights into both the reduced error rate and generalization capability of the detection of face palsy with an accuracy of 99.35%, a recall rate of 99.74%, a precision rate of 99.35%, and an F1-score of 99.54%.

For the skin melanoma detection, in terms of accuracy (92.18%), sensitivity (80.77%), specificity (95.1%), and F1-score (80.84%), the experimental outcomes for the binary classification scenario show a considerable improvement in melanoma

diagnosis. The experimental outcome for multiclass classification showed a significant performance of 89.2% (sensitivity), 96.2% (specificity) for atypical nevus detection, 65.4% (sensitivity), 72.2% (specificity), and for common nevus detection 66% (sensitivity), 77.2% (specificity). In terms of detecting skin melanoma, the proposed classification system outperforms some of the previous methods.

For COVID-19 detection using sound datasets, geometric and photometric augmentation methods on the DeepShufNet model were adopted. Using the Mel COCOA-2 enhanced training datasets, the proposed model detected positive COVID-19 with an accuracy, precision, recall, specificity, and F1-score of 90.1%, 77.1%, 62.7%, 95.98%, and 69.1%, respectively. Comparing the proposed approach to some state-of-the-art techniques, it demonstrated an increased performance.

Finally, it is necessary to establish that the need for more data in classification tasks cannot be overemphasized; however, the presented results of our proposed augmentation methods on deep learning models have shown promising results.

6. GENERAL CONCLUSIONS AND FUTURE RECOMMENDATION

6.1. General Conclusions

In this dissertation, a comprehensive study of the applications of data augmentation techniques has been employed. The application of DAM has aided in improving the learning models by enhancing strong generalization in image classification tasks for small data analysis.

1. The contributions of artificial intelligence methods, especially in medical diseases detection, continues to attract more research attention. However, there are still underlying factors affecting the overall research in this area of small data analytics, especially in computer vison tasks. Some of the leading challenges include the high dependencies on big datasets, feature representation issues as a result of variations in different image properties, the problem of data sparsity, sampling bias, imbalance classes, the likelihood of overfitting of models, thus leading to lower chances of real-life adaptation of the presently existing solutions. The application of data augmentation methods has significantly helped to achieve desired variance of training data, which has resulted in developing an effective learning model and thereby improving the performance results of the classifiers. Some existing challenges of data augmentation methods include the problem with unrealistic augmented images, which has no significant influence in the performance of classification models.

2. This research has created an effective data augmentation method based on the uniqueness of the data used and further analyzed the efficiency of deep learning models in five classification tasks, as summarized below:

    i. The proposed novel DAM method is based on the convolution of the Chebyshev orthogonal functions with the probability distribution functions (PDFs) for cassava disease detection. Four augmentation approaches were employed, namely, motion blur, Gaussian blurring, down-sampling the resolution, and overexposure as strategies to generate synthetic training datasets by reducing the quality of the images. For efficient classification, this study applied a cost-sensitivity and less computation deep neural network model (MobileNetV2).

    ii. This research has introduced a VDRRE image augmentation method based on irregular regions created through Voronoi tessellation and random erasing augmentation for generating synthetic images for facial palsy detection. In addition, few-shot learning based on one-shot and two-shot has been proposed for the intelligent classification of face palsy disease.

    iii. The Covariant Synthetic Minority Oversampling Technique (covSMOTE) augmentation approach has been applied to address the data scarcity and class imbalance problem in skin melanoma detection. The proposed method was based on data oversampling in a nonlinear lower-dimensional embedding manifold for creating augmented images. The DAM is used to produce new augmented images which were used to train the SqueezeNet deep learning model.

iv. For COVID-19 detection using deep breath sounds, an augmentation method based on noise injection and color transformation approach has been proposed. These augmentation strategies were used for generating synthetic training datasets and applied to two feature extraction methods, namely, Mel-spectrogram and GFCC.

3. This research has improved the efficiency, accuracy, recall, precision, and specificity of learning models in small data analysis.

   i. In the face palsy detection, the highest accuracy was achieved by using two-shot learning with the proposed VDRRE method and a hybrid SqueezeNet/ECOC-SVM classifier. Our proposed model achieved 99.34% accuracy with a slight drop in accuracy from 99.34% to 99.07% for one-shot learning. In comparison with state-of-the-art methods, our proposed model outperforms the competition with a great improvement in accuracy, precision, and recall.

   ii. Regarding the cassava disease detection, the accuracy of the baseline network, which was trained on low quality images, achieved 97.7% accuracy. However, the outcome of our proposed model increased by +2.0% after training with the original (high-quality) and augmented dataset (low-quality) images with an accuracy of 99.7%, thus outperforming the currently existing state-of-the-art methods.

   iii. The experimental results in the skin melanoma detection task showed an impressive result in the binary classification case. A significant improvement in the detection of melanoma to an accuracy rate of 92.18% was obtained in the binary detection task, while the best accuracy of 89.2% was obtained in the multiclass classification scenario.

   iv. Regarding the COVID-19 detection task, the application of augmentation by using the noise injection method on Mel-spectrogram and GFCC improved the classification accuracy. The performance evaluation based on accuracy indicates that our proposed augmentation on DeepShufNet gave the best experimental outcome in comparison to the state-of-the-art methods with an accuracy of 90.1%.

4. Further investigation of the augmentation methods efficacy demonstrated significant experimental results through leveraging on the relevance of the domain-specific knowledge to enhance generalization of the model learning process in real-life applications.

6.2. Future Recommendations

Based on the findings of this dissertation on data augmentation techniques and their applications in various classification tasks, the following recommendations for future research are proposed:

1. To explore additional data augmentation techniques, such as advanced geometric transformations, and generative adversarial networks (GANs) for synthetic data generation, and to explore the integration of AI with other emerging technologies, such as augmented reality for data augmentation.

2. To extend the proposed methods to other domains where data scarcity is an issue, such as environmental monitoring, financial modeling, emotion recognition, or social sciences.
3. To perform robust validation of the proposed data augmentation techniques across diverse datasets as well as in real-world scenarios, considering different data characteristics and environmental variables.
4. Given the impact of noisy data on the classifier performance, to develop more advanced methods for noise reduction and handling in data augmentation, particularly in sound and image classification tasks.

7. SANTRAUKA

7.1 ĮVADAS

Dirbtinis internetas tapo sėkminga, greitai augančia ir neįtikėtinai populiaria beveik kiekvienos tyrimų srities dalimi. Taip atsitiko, nes kompiuterinė rega vaidina vis didesnį ir reikšmingesnį vaidmenį ir gali pasiūlyti veiksmingus, mažiau skaičiavimo reikalaujančius, ekonomiškus, tikslius ir kruopščius rezultatus [1, 3]. Vis dėlto viena pagrindinių problemų, su kuriomis susiduriama taikant DI metodus priimant sprendimus, yra duomenų prieinamumas. Paprastai susiduriama su mažų duomenų rinkiniais arba ribotų duomenų rinkiniais [4, 5]. Inžinerinėje srityje mažais duomenimis gali būti vadinami duomenys, turintys mažiau nei 50 pavyzdžių, o akademinių tyrimų atveju – mažiau nei 30 pavyzdžių [6]. Mažų duomenų gali pakakti, jeigu jie tenkina kaupimo, įvairovės, sąžiningumo ir skaidrumo reikalavimus. Priešingai, dauguma mažų duomenų rinkinių nėra pakankamai reikšmingi, kad galima būtų pradėti tyrimą arba sukurti hipotezę. Tačiau, įtraukus kitas įžvalgas ir pastebėjimus, galima sukurti sprendinį arba modelius būsimiems prekių ženklams ar verslams [7]. Šiems duomenų rinkiniams trūksta imperatyvios informacijos ir duomenų analitikos panaudojimo įvairovės, ypač priimant realaus gyvenimo sprendimus / atliekant informuotas užduotis [8].

Duomenų analitiką galima apibūdinti kaip kompiuterinių sistemų naudojimą analizuojant didelių duomenų rinkinius, norint pagrįsti sprendimą. Ši mokslo sritis suvaidino esminį vaidmenį keliose kitose mokslinėse srityse, pavyzdžiui, atpažįstant modelius, tiriant operacijas, skaičiuojamojo intelekto, mašininio mokymosi atvejais ir pan. [270]. Nesenais tyrimais nustatyta, kad mažų duomenų analitika yra tokia pat svarbi, kaip ir didelių duomenų analitika, ypač priimant realaus laiko arba informuotus sprendimus. Ankstesnių vaizdų apdorojimo tyrimų metu buvo pabrėžta, kad, naudojant geresnės kokybės mažų, atrinktų duomenų rinkinius, galima gauti geresnius rezultatus nei dirbant su prastos kokybės didelių duomenų pavyzdžiais [9]. Vis dėlto, kai duomenų analitika taikoma realiame pasaulyje tokiais skirtingais tikslais, kaip ligos diagnozavimas ir užduočių klasifikavimas, reikia pakankamų ir subalansuotų duomenų rinkinių, kad jie būtų veiksmingai panaudojami dirbant su DI. Deja, kai mažų duomenų rinkiniai ir dirbtinio intelekto metodai taikomi tokiose srityse, kaip sveikatos priežiūra, augalų ligos, garso aptikimas, gamyba ir pan., vis dar susiduriama su paklaidomis, skaičiavimo sudėtingumu, nesubalansuotais duomenų rinkiniais, prastu apibendrinimu ir pan.

Be to, dėl kelių pagrįstų veiksnių, turinčių įtakos duomenų surinkimui, mažų duomenų rinkiniai medicinos ir sveikatos srityje kelia didelių iššūkių, pavyzdžiui, duomenų privatumas grindžiamas jautriais / gyvybiniais pacientų duomenimis (asmeniniai, konfidencialūs duomenys), atsiranda išlaidų, taikomi darbuotojų ir kompetencijos reikalavimai aiškinimui, pacientų duomenų anonimizuojami prieš juos naudojant ar jais dalinantis [10]. Kitos mažų duomenų rinkinių taikymo DI problemos – tai didelis variantiškumas, dėl ko rezultatai išsikraipo, arba duomenimis grįsti modeliai yra pernelyg apibendrinami [4]. Taigi, reikia skubiai atlikti tyrimus, kaip sukurti veiksmingus modelius, galinčius išspręsti mažų duomenų problemas.

Atsižvelgiant į neseną dirbtinių modelių sėkmę, pristatomą literatūroje, taikomi duomenų papildymo metodai padėjo spręsti mažų duomenų problemas, sukuriant virtualius / dirbtinius pavyzdžius, grindžiamus ankstesnėmis žiniomis, gautomis iš konkrečių ribotų mokymo duomenų [11]. Duomenų papildymas (DP), naudojamas kaip priemonė naujiems duomenų rinkiniams sudaryti, pademonstravo daug žadančius rezultatus ir labai pagerino mokymosi klasifikatorių veikimą tokiose srityse, kaip vaizdų atpažinimas, signalo ir balso atpažinimas, pramoniniai sektoriai [6], objekto aptikimas ir pan. Kitaip sakant, vienas reikšmingiausių metodų, kaip veiksmingai pagerinti klasifikatorių darbą su mažų duomenų pavyzdžiu, yra pritaikyti duomenų papildymo principą [12].

Duomenų papildymo metodas (DPM) – tai gerai žinomas ir priimtinas metodas, taikomas norint padidinti esamų mokymo duomenų įvairovę ir kiekį, tiesiogiai nerenkant papildomų duomenų pavyzdžių [13]. DPM reikšmingas ne vien dėl to, kad padidina duomenų variantiškumą, bet ir todėl, kad sumažina disbalansą, duomenų rinkinių paklaidas ir padeda išvengti perpildymo per modelius. Todėl jis veiksmingai taikomas apdorojant vaizdus ir atliekant kompiuterinės regos užduotis, siekiant susintetinti ir subalansuoti mokymo duomenų rinkinius, sukuriant papildomus mažumos klasių pavyzdžius [14, 15]. Duomenų papildymo metodas daro įtaką ir kitoms sritims, kuriose nepakankami duomenys irgi kėlė didelių problemų, pavyzdžiui, geoterminėje srityje, nuspėjant didelės entropijos lydinių (DEL) kietumą, vertinant klientų kreditingumą ir pan. Šis metodas būtinas, norint padidinti mokymo duomenis ir taip pasiekti patenkinamų rezultatų bei išspręsti nepakankamų duomenų klausimą [16]. Neseniai naudojant DPM buvo sukurti dirbtiniai duomenys, kas smarkiai patobulino kompiuterinės regos tyrimus – tai aiškiai parodyta didelės apimties vizualinio vaizdų atpažinimo konkurse (ILSVRC) [17]. Ankstesnėje literatūroje naudoti DPM pavyzdžiai, kaip tobulinamos regos užduotys: vaizdų apkarpymas, apvertimas (horizontalus ir vertikalus), didinimas ir mažinimas, spalvų derinimas, Gauso suliejimas, sukimas, triukšmas ir pan. [15, 16, 18]. Be to, įrodyta, kad DP metodai padeda gauti žinių iš iškreiptų duomenų rinkinių. Vis dėlto, esant dideliam disbalansui mažų duomenų rinkinyje, duomenų papildymo metodai gali nepadėti pasiekti laukiamų skirtumų išskirtiniais duomenų rinkinių perbalansavimo atvejais [14], tačiau jie vis tik padės pagerinti klasifikatorių veiklą. Atliekant nesenus tyrimus, stebėti skirtingi duomenų papildymo metodai, taikomi nepakankamų mokymo duomenų problemai spręsti, ir nustatyta, kad populiariausi metodai yra tradiciniai arba geometriniai metodai, pavyzdžiui, atsitiktinis sukimas, apvertimas, transformavimas ir pan. Kiti pažangūs papildymo metodai, aptinkami duomenų transformavimo literatūroje, yra generatyviniai priešiški tinklai (GPT), variacinis autoenkoderis (VAE) ir pan., kurių tikslas – pakeisti rezultatus, kuriant netikrus vaizdus [19]. Daugumoje ankstesnių tyrimų rodomi progresuojantys kai kurių duomenų papildymo metodų taikymo rezultatai, tačiau kai kuriems esamiems metodams vis dar trūksta gebėjimo apibendrinti. Taip yra dėl triukšmingo fono, problemų dėl netinkamo aiškinimo ar nesuderinamo žymėjimo [20]. Be to, DI metodų veikimo tobulėjimas visiškai priklauso nuo esamų duomenų dydžio, kokybės ir įvairovės [4].

7.1.2. Darbo objektas

Šio tyrimo objektas yra duomenų papildymo metodų tyrimas ir tobulinimas, naudojant veiksmingus gilaus mokymosi modelius mažų duomenų analizei. Skirtingi padidinimo metodai pateikiami ir taikomi įvairiems ligos aptikimo atvejams naudojant tiek vaizdų, tiek garso duomenų rinkinius. Šiame darbe sąvoka „maži duomenys" apibrėžiama kaip santykinai nepakankamas duomenų dydis, kuriam būdinga ribota imtis ir įvairovė.

7.1.3. Darbo tikslas

Tyrimo tikslas – sukurti patobulintus duomenų papildymo metodus, leidžiančius geriau apibendrinti ir modeliuoti mažus duomenų rinkiniais, skirtus vaizdo apdorojimui ir garso atpažinimo sistemoms.

7.1.4. Darbo uždaviniai

Siekiant darbo tikslo, nustatyti šie uždaviniai:
1. Peržvelgti ir analizuoti literatūrą apie esamų duomenų papildymo metodus, siekiant nustatyti veiksmingus mažų duomenų rinkinių vertinimo metodus klasifikuojant ligas.
2. Supažindinti ir pritaikyti sukurtus išplėtimo metodus vaizdo ir garso duomenų rinkinių etalonams, kuriais siekiama išspręsti mažas duomenų problemas ir tobulinti.
3. Pasiūlyti ir įvertinti modifikuotus išplėtimo metodus, kurie pagerina ligos aptikimą ir taikyti šiuos metodus vaizdo ir garso duomenų rinkiniams.
4. Ištirti ir palyginti papildomus metodus konkrečiose programose srityse ir įvertinti našumą bei prisitaikymą prie realaus scenarijaus.

7.1.5. Praktinė vertė

Praktinė keturių uždavinių ir sukurtų metodų, aprašytų šioje disertacijoje, vertė yra reikšminga įvairiose realaus gyvenimo srityse:
1. Nauji duomenų papildymo metodai, naudojami manioko (*'cassava'*) ligai nustatyti, gali būti praktiškai pritaikyti žemės ūkyje. Naudojant vaizdų praplėtimo metodus, tokius kaip suliejimas ir per didelis eksponavimas, modelis gali tiksliau nustatyti augalų ligas įvairiomis lauko sąlygomis. Šis taikymas gali leisti veiksmingiau valdyti augalų ligas, padidinti derlių ir kurti naudą ūkininkams bei nuo žemės ūkio gamybos priklausomoms ekonomikoms. Be to, ateityje šis metodas gali būti pritaikytas tokiose srityse, kaip pramoninės kokybės kontrolė. Gamyboje, pavyzdžiui, automobilių ar elektronikos pramonėje, šis metodas galėtų būti naudojamas gaminių defektams aptikti esant įvairioms apšvietimo ir aplinkos sąlygoms, padidinant automatinių kokybės įvertinimo sistemų tikslumą.
2. VDRRE vaizdų praplėtimo metodas veido paralyžiui aptikti gali turėti svarbių rezultatų sveikatos priežiūrai. Gebėjimas tiksliai aptikti ir klasifikuoti veido paralyžių iš vaizdų gali supaprastinti diagnostiką, kad būtų galima greičiau ir

veiksmingiau gydyti. Šis metodas gali būti ypač vertingas atokiose arba nepakankamų išteklių srityse, kur specialistų prieinamumas yra ribotas. Be to, ateityje šis metodas galėtų būti taikomas kuriant pažangias veido vaizdais pagrįstas emocijų atpažinimo sistemas, kad galėtume geriau suprasti žmogaus emocijas ir reakcijas realiuoju laiku.

3. CovSMOTE praplėtimo technika odos melanomai aptikti tenkina esminį dermatologijos ir onkologijos poreikį. Duomenų praplėtimas SqueezeNet modeliu pagerina odos melanomų aptikimą ir klasifikavimą, o tai yra labai svarbu ankstyvai diagnostikai ir gydymui. Šis metodas galėtų būti integruotas į nuotolinės medicinos platformas, padedančias atlikti odos patikrinimą. Tai gali sumažinti sveikatos priežiūros sistemų naštą. Be to, ateityje covSMOTE metodas, kuris yra orientuotas į klasių disbalanso problemą, taip pat galėtų būti pritaikytas kitoms reikmėms, pavyzdžiui, žemės ūkyje kenkėjams aptikti. Išmokius modelius atpažinti nedidelius, bet kritinius pokyčius esant kenkėjams, šis metodas gali padėti anksti aptikti ir valdyti pasėlių užkrėtimo atvejus. Tai gali padėti išvengti didelio masto žemės ūkio veiklos nuostolių.

4. Išplėtimo metodas, apimantis triukšmo įterpimo ir vaizdo transformavimo metodus, skirtus COVID-19 prognozuoti, gali turėti teigiamą poveikį pandemijos aptikimui bei valdymui ir visuomenės sveikatos pasekmėms. Dėl galimybės aptikti COVID-19 iš kvėpavimo garsų naudojant AI gali būti sukurti neinvaziniai greitojo testavimo metodai. Tai gali palengvinti didelio masto patikrinimą, ypač tose vietose, kur tradiciniai testavimo metodai yra logistiškai sunkūs. Be to, ateityje šis metodas gali būti naudojamas ir kitoms reikmėms, pavyzdžiui, kaip aplinkos stebėjimas, ypač nustatant ir analizuojant garsų taršą natūralioje aplinkoje arba gyvūnų garso duomenų rinkiniuose, taip prisidedant prie biologinės įvairovės tyrimų ir aplinkos apsaugos pastangų.

Šie metodai parodo, kaip pažangūs duomenų papildymo būdai gali smarkiai pagerinti DI modelio veikimą įvairiose srityse – nuo žemės ūkio iki sveikatos apsaugos. Pagerinus aptikimo ir diagnostikos tikslumą bei efektyvumą, šie metodai gali turėti teigiamą poveikį įvairiems sektoriams ir prisidėti prie teigiamų socialinių rezultatų. Panaudodami ir pritaikydami šias pažangias technologijas, galime išspręsti daugybę sudėtingų problemų įvairiose tarpusavyje nesusijusiose srityse. Tai demonstruoja didelį DI ir duomenų papildymo sektorių transformacinį potencialą.

7.1.6. Darbo teiginiai

Įrodyta, kad duomenų papildymo metodai yra labai svarbi taktika didinant dirbtinio intelekto modelių galimybes keliose srityse. Pakeitus ir padidinus mokymo duomenų rinkinį, šis metodas pagerina jo dydį, įvairovę ir atsparumą, o tai savo ruožtu padidina mašininio mokymosi modelių efektyvumą atliekant sudėtingas užduotis. Vaizdų klasifikavimo srityje duomenų papildymo metodą galima naudoti norint pagerinti mokymo duomenų dydį ir mokymo variantiškumą.

1. Voronojaus dekompozicijos atsitiktinio regiono trynimas (VDRRE) taiko atsitiktine tvarka netaisyklingą uždengimą, pagaunant įvairias veido išraiškas ir unikalias variacijas veido paralyžiumi sergančių asmenų vaizduose. Be to,

vaizdo spalvų histogramos išlyginimo ir vaizdo kokybės sumažinimo metodų derinys sudaro hibridinį praplėtimo metodą, kuris įveda įvairovę manioko (*cassava*) lapų ligos vaizduose. Ši simuliacija apima skirtingas apšvietimo sąlygas, vaizdo kokybę ir triukšmo lygį manioko (*cassava*) ligos atpažinimo užduotyje. Be to, CovSMOTE metodo taikymas palengvina sintetinių duomenų generavimą mažumos klasei (odos melanoma), atsižvelgiant į duomenų kovariacijos struktūrą. Šie praplėtimo būdai sukuria didesnius ir daugiau įvairovės pavyzdžių, turinčius rinkinius, efektyviai pagerina giluminio mokymosi modelius apibendrinant nematytus vaizdus ir tiksliai klasifikuojant.

2. Garso signalų apdorojimo srityje garso signalų konvertavimas į spektrogramų vaizdus leidžia pritaikyti vaizdų praplėtimo metodus, skirtus mokymo duomenų įvairovei padidinti. Nuodugnesnis ir įvairesnis duomenų rinkinys sukuriamas naudojant tokias technologijas, kaip garso aukščio poslinkis, ištempimas laike, triukšmo pridėjimas ir spektrogramos savybių keitimas. Tai sustiprina duomenų rinkinį, pagerina modelio mokymo efektyvumą, jam leidžiant nustatyti ir tiksliau suskirstyti garso signalus į klases esant įvairioms aplinkoms ir sąlygoms.

7.1.7. Mokslinis patvirtinimas

Visi šio darbo rezultatai unikalūs ir pateikiami iš viso devyniuose leidiniuose. Informatikos, mašininio mokymosi ir elektronikos srityse yra šeši tarptautinių mokslinių žurnalų straipsniai pavadinimu „ISI Mokslo tinklas" su nuorodų rodykle ir trys žemiau išvardinti konferencijų medžiagos leidiniai. Visą leidinių sąrašą galite rasti skyriuje „OLUSOLA OLUWAKEMI ABAYOMI-ALLI LEIDINIŲ DISERTACIJOS TEMA SĄRAŠAS".

1. Data Augmentation and Deep Learning Methods in Sound Classification: A Systematic Review. Electronics, 11(22), 3795.
2. Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques. Electronics, 11(16), 2520.
3. An ensemble learning model for COVID-19 detection from blood test samples. Sensors, 22(6), 2224.
4. Cassava disease recognition from low-quality images using enhanced data augmentation models and deep learning. Expert Systems, 38(7), e12746.
5. Few-shot learning with a novel Voronoi tessellation-based image augmentation method for facial palsy detection. Electronics, 10(8), 978.
6. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifolds. Turkish Journal of Electrical Engineering and Computer Sciences, 29(8), 2600-2614.
7. BiLSTM with data augmentation using interpolation methods to improve early detection of Parkinson's disease. In 2020 15th Conference on Computer Science and Information Systems (FedCSIS). pp. 371-380. IEEE.

8. Data augmentation using principal component resampling for image recognition by deep learning. In Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19 (pp. 39-48). Springer International Publishing.

9. Abayomi-Alli, O. O., Sidekerskienė, T., Damaševičius, R., Siłka, J., & Połap, D. (2020). Empirical Mode Decomposition Based Data Augmentation for Time Series Prediction Using NARX Network. In Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part I 19 (pp. 702-711). Springer International Publishing.

7.1.8. Disertacijos mokslinis inovatyvumas

Disertacijos mokslinis inovatyvumas yra galimybė pagerinti dirbtinį intelektą sprendžiant bei analizuojant mažus duomenų kiekius turinčius uždavinius ir gerinti giluminio mokymosi modelių gaunamus rezultatus taikant naujoviškus duomenų papildymo metodus. Ši mokslinė disertacija nagrinėja kelias duomenų papildymo metodų naujoves, skirtas skirtingiems vaizdams su unikaliais duomenimis apdoroti, analizės metu akcentuojant vaizdo ypatybes, susijusias su nustatytomis problemomis.

Pirmiausia buvo pristatytas naujas duomenų padidinimo metodas, pagrįstas Voronojaus dekompozicijos atsitiktinio regiono trynimu. Voronojaus metodas buvo pritaikytas sintetinių duomenų rinkinių variacijoms generuoti. Nustatytas geresnis aptikimo tikslumas (99,34%) nei taikant šiuo metu esamus padidinimo metodus su GAN. Tikslumo rodiklis padidėjo +4,53 %.

Antra, siekiant pagerinti manioko (*cassava*) ligos aptikimą, buvo naudojama Čebyševo stačiakampių funkcijų konvoliucija su tikimybių pasiskirstymo funkcijomis (PDF) ir vaizdo kokybės mažinimo didinimo metodu. Eksperimentai parodė, kad žemos kokybės vaizdų tikslumas buvo 97,7 %, o +2,0 % tikslumo prieaugis buvo gautas naudojant originalius (aukštos kokybės) su papildytais duomenų rinkiniais (žemos kokybės) vaizdus, kurių bendras tikslumas buvo 99,7 %. Siūlomų papildymo metodų rezultatai pasirodė geresni, palyginti su kai kuriais jau esamais metodais, naudojant analizuojamus duomenų rinkinius.

Trečia, ištyrus siūlomą CovSMOTE padidinimo metodą odos melanomai aptikti, buvo pasiektas 92,18 % aptikimo tikslumas.

Ketvirta, disertacijos rengimo metu buvo ištirtas dar vienas duomenų papildymo metodas, kuris nagrinėjo spalvų / fotometrinių ir triukšmo įpurškimo įtaką garso duomenų rinkiniams. Šie papildymai buvo pritaikyti dviem funkcijų išgavimo metodams: Mel spektrogramai ir GFCC. Aptikimo tikslumas pasirodė greitesnis, palyginti su esamais moderniausais metodais, kurių tikslumas yra 90,1 %.

7.1.9. Disertacijos struktūra

Disertaciją sudaro šeši skyriai urie yra taip: pirmasis skyrius yra įvadas, kur trumpai apžvelgiamas tyrimo originalumas, išdėstomi tikslai ir uždaviniai. Antrame skyruje pateikiama išsami dirbtinio intelekto metodų apžvalga, detaliai išanalizuoti

smulkių duomenų analizės ir duomenų papildymo metodai. Šiame skyriuje taip pat aprašoma literatūros apžvalga ir įvairios duomenų papildymo taikymo sritys mažų duomenų analizėje, siekiant supažindinti skaitytojus su terminais, kurie bus naudojami tolesniuose skyriuose, ir taikymo sritimis. Trečias skyrius yra įvairių viešai prieinamų mažų duomenų rinkinių, naudojamų šioje disertacijoje, santrauka. Ketvirtame skyriuje aprašyta siūloma metodika, pristatyta siūlomų didinimo metodų architektūra ir teorinis pagrindas.

Penktas skyrius pristato gautus eksperimentų rezultatus ir detaliai palygina siūlomų metodų rezultatus su jau egzistuojančiais moderniausias metodais. Galiausiai šeštame skyriuje apibendrinamos išvados ir pateikiamos rekomendacijos ateities darbams.

7.2 LITERATŪROS APŽVALGA

Duomenų papildymas – tai žingsnis prieš mašininį mokymąsi ir gilų mokymąsi, kai, siekiant sukurti naujus mokymo pavyzdžius, pakeičiamas ir išplečiamas pradinis duomenų rinkinys [52]. Šis metodas dažnai naudojamas su mažų duomenų rinkiniais, siekiant pagerinti mokymo duomenų kokybę ir išvengti perpildymo problemos. [53] apibendrinamas duomenų papildymo tikslas kaip galimybė apibendrinti, nesumažinant modelių reprezentacinės galios ir nepakeičiant kitų hiperparametrų. Taigi, duomenų papildymo metodus galima taikyti skirtingai, pavyzdžiui, linijinei ar nelinijinei transformacijai, papildomų kintamųjų pridėjimui ir duomenų sukūrimui, naudojant generatyvinius modelius [54]. [55] aprašoma, kad duomenų papildymas sistemingai deformuoja neapdorotus įvesties duomenis, kad galima būtų smarkiai padidinti mokymo duomenų rinkinio kiekį. Duomenų papildymas taikomas skirtingose srityse, pavyzdžiui, NLP [56, 57], kalbos atpažinimo [58], veido atpažinimo [59], judesio aptikimo [21], ligos diagnozavimo [60] ir pan. Pagrindinis duomenų papildymo tikslas – padidinti mokymo duomenų rinkinio kiekį ir diversifikaciją, sumažinti perpildymą ir padidinti modelio tvirtumą ir generalizaciją. Papildyti duomenys naudojami ML modeliams mokyti, pateikia papildomus mokymo pavyzdžius ir padeda padidinti modelio tikslumą. Be to, taip pagerinamas bendras gilaus mokymosi modelių veiksmingumas, bendras stabilumas, mokymo rezultatų standartas [61] ir „mažų duomenų problemos" sprendimas [28]. Norint maksimizuoti gilaus mokymosi modelių pagrindimą mažais duomenimis, buvo sukurta daugiau praktinių sprendimų, pavyzdžiui, pašalinimo sureguliavimas, perkeliamas mokymasis, išankstinis mokymas ir grupės normalizavimas [56, 62, 63]. Vaizdai gali būti papildomi įvairiais tradiciniais arba atsitiktinės transformacijos metodais, tiksliau, naudojant didinimą ir mažinimą, sukimą, apvertimą [64], pločio pakeitimą, aukščio pakeitimą, atsitiktinį apkarpymą [65], nukirtimą, mastelio keitimą, atsitiktinį ištrynimą [66] ar pašalinimo sureguliavimą [18, 67], pridedant triukšmo (druska ir pipirai, Gauso triukšmas, Poissono triukšmas), ir pavyzdžių sintezę (naudojant generatyvinius modelius), elastinę deformaciją [68] ir parametrų išplėtimą [69]. Kiti įdomūs metodai [68]: normalizuotas atsitiktinio poslinkio laukas, užblokavimo metodas [70], iškarpa [71], nugludinimo kaukės filtras (USM) [67], SMOTE [72].

Neseniai pradėta intensyviai tirti, kaip DPM gali būti naudojamas kalbai atpažinti. Čia pagrindinis dėmesys skiriamas skirtingiems garso pakeitimo būdams. Ankstesnėse studijose siūlytų metodų pavyzdžiai buvo vokalinio trakto ilgio normalizavimas (VTLN), greičio trikdymas, specialus papildymas [58], duomenų iškreipimas [68]. Dar viena įdomi sritis, kurioje naudojamas DPM, yra sustiprintas mokymasis. [73] pristatomas K-mišinys, naudojant Koopmano nekintamą poerdvę mokymo pavyzdžio dydžiui padidinti. Trukmės modifikavimo principą mokymo duomenims papildyti, siekiant aptikti dizartrijos sutrikimą, pasiūlė [74]. Klasifikuojant garsus, prie mokymo duomenų įvairovės ir variacijų prisidėjo ir kiti metodai, pavyzdžiui, foninio garso pridėjimas, patalpos impulso atsakas, garso aukščio ir laiko pasikeitimas ir pan. Taikomi ir tokie pažangūs papildymo metodai, kaip Conv GAN [75], Wasserstein GAN [76]. Puikios DPM metodų galimybės garso ir audioklasifikavimo srityje padėjo klasifikuoti ir pagerino bendras kompetencijas daugelyje kitų sričių. Vis dėlto, kai DPM metodai naudojami su mažų duomenų rinkiniais, susiduriama su nedideliu nukrypimu klasifikuojant rimtumą, kas paveikia mokymosi modelio veikimą.

7.3. DUOMENŲ RINKIMAS

Buvo išanalizuoti ir šioje disertacijoje panaudoti trys duomenų tipai. Tai vaizdai, garsas ir skaitmeniniai duomenys (biomedicininiai įrašai). Duomenų rinkiniai yra maži, ir susiduriama su nesubalansuotų klasių problema. Pirmieji šiame tyrime panaudoti vaizdų duomenys – tai YFP duomenų rinkinys, skirtas veido paralyžiui diagnozuoti. Jame tėra tik Belo paralyžiaus vaizdo įrašai [99], kurie konvertuoti į 1105 veido paralyžiaus vaizdų seką. Sveikų ir su veido paralyžiumi nesusijusių duomenų rinkiniai gauti iš Kalifornijos technologijos instituto duomenų bazės [223], kurią sudaro 27 skirtingų asmenų 450 veido vaizdai. Antrasis duomenų rinkinys buvo susijęs su maniokų ligų diagnozavimu. Panaudotas „iCassava" duomenų rinkinys [108], kuriame yra 9436 duomenų vienetai. Diagnozuojant odos melanomą, naudoti PH2 duomenų rinkiniai su 200 vaizdų, suskirstytų į tokias kategorijas: 80 įprastų apgamų, 80 netipiškų apgamų ir 40 melanomų [141]. Galiausiai, tyrimui naudotas ir antrinis duomenų tipas. „Coswara" duomenų rinkinys [240] naudotas garsui klasifikuoti. Jį sudaro 2130 įrašų iš devynių skirtingų kategorijų ir septynių klasių. Visi šie duomenų rinkiniai – tai kitų tyrėjų anksčiau atliktų tyrimų, paskelbtų straipsniuose, tęsimas.

7.4. TYRIMO DIZAINAS IR METODAI

7.4.1. Siūlomas Voronojaus skaidymu grindžiamas atsitiktinės srities ištrynimas (VDRRE) – veido paralyžiaus diagnozavimo metodas

Visais atvejais vieno kadro mokymosi pavyzdys buvo sustiprintas atsitiktinio atskyrimo trynimo vaizdų papildymu. Buvo pasiūlytas VDRRE metodas ir ištirtas jo naujumas stiprinant vaizdus. 2D plokštumos vaizdai buvo padalinti į vieną šalia kitos esančias konkrečių taškų zonas, o šių taškų koordinatės gaunamos naudojant sutartinius skaičius, paimtus iš vienodo paskirstymo. Iš esmės vaizdas atskiriamas parenkant N atsitiktinai pasiskirsčiusių taškų, kas sudaro Voronojaus mozaiką. Taigi, vaizdai dalijami Voronojaus zonoje aplink kiekvieną konkretų objektų rinkinį [249].

Pavyzdžiui, generatorių rinkinys $P = \{p_1, p_2, ... p_n\} \in R^2$. Atstumas iš bet kurios X zonos plokštumoje iki generatoriaus taško P_n žymimas $dist(X, p_n)$. Pasirinkime artimiausią generatorių $p_n \, \epsilon P$ su konkrečiu metriniu $dist$ visų galimų X vietų S atžvilgiu. Jeigu pasirodys arti dviejų generatorių P, intervalas tampa kraštu; kita vertus, jeigu yra arti daugiau nei dviejų generatorių, padėtis pasikeičia į viršūnę. Tegul $dist\,(p, X_i)$ reiškia euklidinį atstumą tarp bet kurios padėties p erdvėje ir pradinės zonos X_i. (7.1) lygtimi galima apskaičiuoti pusiaukampinę tarp X_i ir X_j, o (7.2) lygtimi gali būti nustatoma sritis, kur X_i dominuoja X_j:

$$b(X_i, X_j) = \{dist(p, X_i) = dist(p, X_j)\}, \qquad (7.1)$$
$$Dom(X_i, X_j) = \{dist(p, X_i) \leq dist(p, X_j)\}, \qquad (7.2)$$

čia $b(X_i, X_j)\ yra$ statmena linijos, jungiančios X_i, X_j, pusiaukampinė

$Dom(X_i, X_j)$ yra X_i vyravimo X_j atžvilgiu, sritis

kur $i \neq j$, o $dist(p, X_i)$ yra svertinis atstumas tarp p ir X_i. Paprasto X_i Voronajaus sritį galima apibrėžti taip:

$$V(X_i) = \cap_{X_j \in X \setminus \{X_i\}} Dom(X_i, X_j), \qquad (7.3)$$

$V(X_i)$ vadinama svertine Voronojaus sritimi.

Atliekant šią užduotį, vaizdas išskaidytas į keletą Voronojaus elementų su vaizdo ribomis. Šis procesas vadinamas Voronojaus mozaika. Galiausiai, siekiant uždarumo ir norint sukurti naują vaizdą, netaisyklingas pasirinktas Voronojaus elementas prisipildo atsitiktinai parinktų pikselių spalvų verčių, gautų iš tolygaus pasiskirstymo, vaizduojamo 7.1 pav.
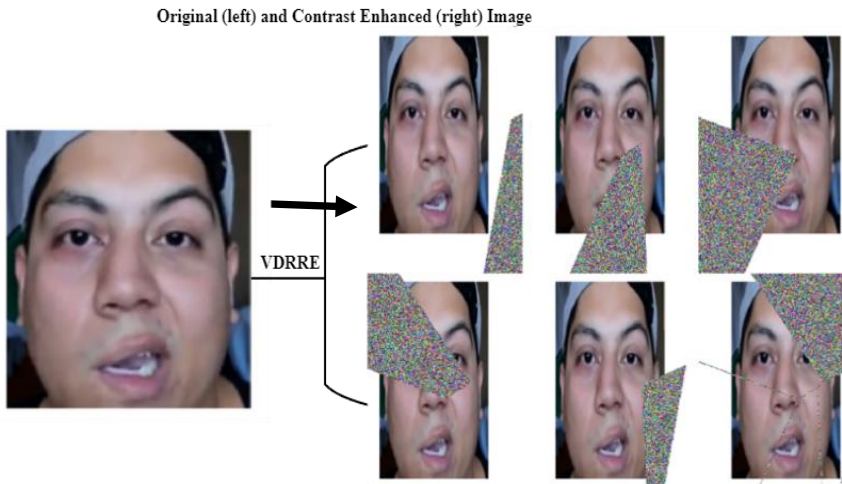
7.1 pav. Originalus vaizdas (kairėje), papildytas vaizdas, gautas taikant siūlomą atsitiktinės srities ištrynimo metodą (dešinėje).
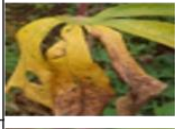
7.4.2. Vaizdo spalvų histogramų sulyginimo metodas – maniokų ligų atpažinimas

Norint aprašyti bet kokią funkciją funkcijos erdvėje, galima naudoti pagrindinių funkcijų, vadinamų stačiakampėmis funkcijomis, rinkinį. Vertinant pagal rezultatus, anksčiau šie algoritmai nebuvo naudojami histogramoms sulyginti arba duomenims papildyti sprendžiant klasifikavimo ir vaizdų apdorojimo problemas. Čebyševo daugianariai – tai unikalios stačiakampių ultrasferinių daugianarių variacijos. Vertinant, kad gauta funkcija iš viso verta 1, transformacija įvykdoma naudojant Čebyševo daugianarį su atitinkama PDF verte. Atskiros neapdoroto vaizdo RGB kanalų histogramos dėl šios transformacijos pakeičiamos. Reikia nepamiršti, kad skirtingos stačiakampės funkcijos gali būti naudojamos įvairiems kanalams. Be to, vaizdas transformuojamas kiekviename pikselyje, taip sukuriant dirbtinį vaizdą. Fig. 4.12 (žr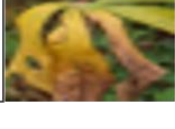. pagrindinėje disertacijos dalyje) pateikiama PDF transformacijos iliustracija, kai galutinis elementas paliekamas tuščias.

Todėl, norint sumažinti testuojamų vaizdų kokybę, buvo sukurti duomenų papildymo metodai. Taip pavyko gauti prastesnės kokybės vaizdus, kurie suskirstyti pagal keturis transformacijos metodus. Kokybės sumažinimo laipsnis vertintas kaip mažas, vidutinis ir didelis kokybės sumažinimas. Rezultatai pateikiami 7.1 lentelėje.

   a. Skiriamosios gebos mažinimo pavyzdžiai: neapdorotas vaizdas, sumažintas iki $32 \times 32; 64 \times 64; 96 \times 96; 128 \times 128; 160 \times 160; 192 \times 192$ pikselių ir padidintas iki 224×224 pikselių kaip galutinio reikalingo dydžio nerviniam tinklui.
   b. Gauso suliejimas: norint supinti neapdorotus įvedamus vaizdus, buvo naudotas Gauso filtras su standartiniu 1, 1.5, 2, 2.5 ir 3 nukrypimu.
   c. Judesio suliejimas: siekiant nustatyti judesio kampą iš (0,360) kampų spektro, buvo naudotas vienodo pasiskirstymo atsitiktinis skaičius, o norint įvertinti kameros linijinį judėjimą, naudotos 10, 15, 20, 25 ir 30 pikselių reikšmės.
   d. Perlaikymas: vaizdų kanalų amplitudės padidintos 1.25 (1.5, 1.75, 2 ir 2.5) faktoriumi, taip atkartojant perlaikytų nuotraukų sąlygas.

7.1 Lentelė. Papildytų vaizdų rezultatai, taikant IQRM.

Methods/ Level	Resolution downsampling	Gaussian Blur	Motion-Blur	Overexposure
Low = 0				
Medium = 1				
High = 2				

7.4.3. Siūlomas CovSMOTE argumentavimo metodas odos melanomai diagnozuoti

Duomenims papildyti buvo pritaikytas modifikuotas dirbtinio mažumos įrašų kūrimo metodas (SMOTE) [153]. Taikant SMOTE, gaunami duomenys iš mažumos klasės dirbtinių įrašų. SMOTE didžiausią dėmesį sutelkia į mažumos klasės kaimynų, artimų mažumos klasei, atpažinimą. Taškas tarp neapdoroto pavyzdžio ir kaimyninio pavyzdžio pasirenkamas atsitiktinai. Buvo taikomas Cov-SMOTE [155] metodas. Tai yra modifikuotas SMOTE, į kurį įtraukiama kovariacijos matrica, skirta priklausomybės ryšiams tarp savybių nustatyti. Įvertinus kovariacijos matricą, gauti nauji arba pakaitiniai pavyzdžiai, turintys padėti atrasti pusiausvyrą tarp abiejų klasių (mažumos ir daugumos). Algoritmas veiksmingai užtikrina, kad visi dirbtiniai kovariacijos pavyzdžiai būtų reikalingi kiekvienos savybės intervale, kaip vaizduojama 7.2 pav. ir 7.3 pav. rodomas PH2 duomenų rinkinio dauginimo erdvės rezultatas. Panašus metodas naudotas ir gaminant naujus mažumos klasės pavyzdžius mažesnių matmenų dauginimo erdvėje. Gavus naujus pavyzdžius, labai svarbus tampa virsmas iš dauginimo erdvės į vaizdų erdvę. Daugianarės regresijos metodas taikytas dėl linijiškumo dauginimo erdvėje nebuvimo, todėl jis padėjo gauti geriausią virsmą dirbtiniame kaimyne. Norint gautų žymėjimą iš dauginimo erdvės į spalvų erdvę, buvo atlikta kubinė daugianarė regresija. Gauti rezultatai (dirbtiniai vaizdai) pateikiami 7.4 paveiksle.
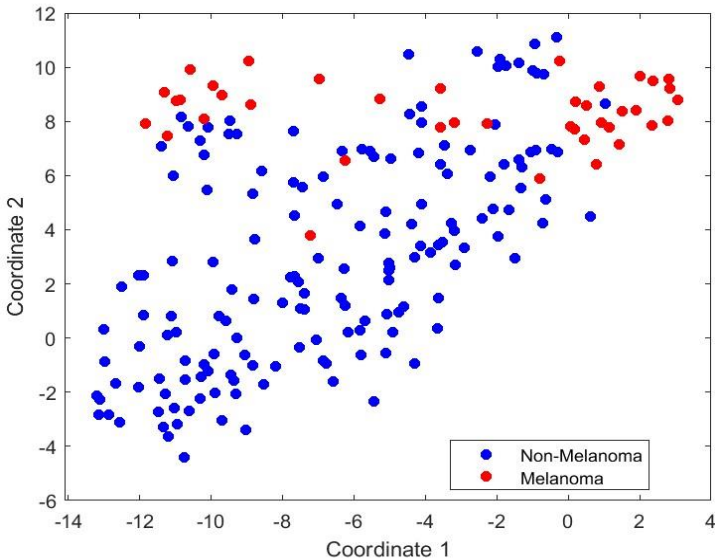
7.2 pav. PH2 duomenų rinkinio klasių vizualizacija
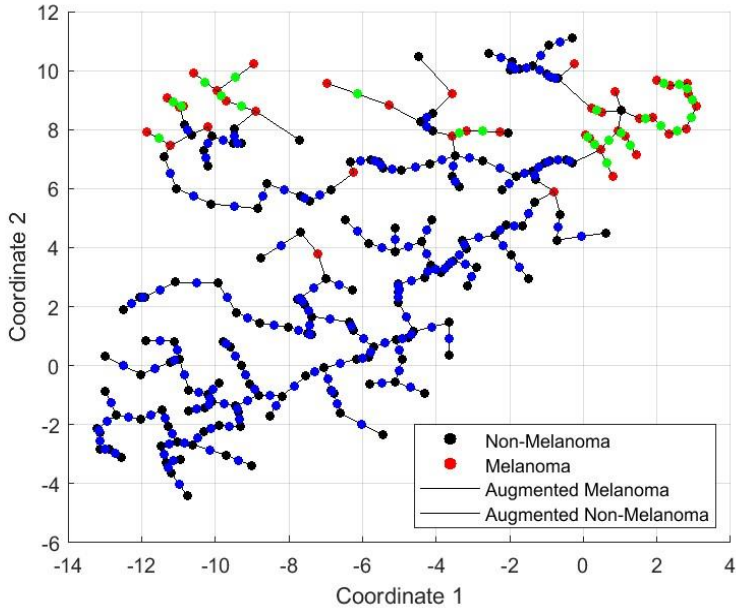
7.3 pav. PH2 duomenų rinkinio klasių vizualizacija; (b) dirbtinių pavyzdžių, sudarytų kaip nauji atvejai (žymimi žaliais taškeliais) atvaizdavimas.
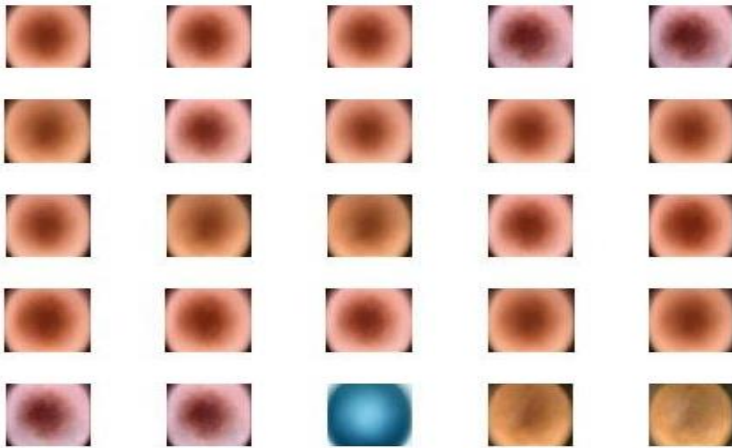
7.4 pav. Papildytų vaizdų, sukurtų taikant CovSMOTE metodą, pavyzdžiai

7.4.4. Siūlomi garso papildymo metodai (fotometrija ir triukšmo įvedimas)

Garso įrašai paversti vaizdais, naudojant du savybių ištraukimo metodus, tiksliau – Mel spektrogramas ir gama tono dažnio Furjė kosinuso koeficientus (GFCC), kaip vaizduojama 7.5 ir 7.6 paveiksluose. Vėliau, siekiant sustiprinti mokymo vaizdų mažumos klasę, panaudotos dvi duomenų papildymo metodų rūšys

131

ir sukurtas naujas dirbtinis duomenų rinkinys. Dirbtinių duomenų sukūrimui ir mokymo duomenų rinkiniui pagerinti naudoti fotometrinis, arba spalvų transformacijos, ir triukšmo įvedimo metodai.

Dirbtinių arba papildytų duomenų, vadinamų COswara-COvid papildytais duomenų rinkiniais, kartu vadinamais COCOA, aprašymas:

1. Fotometrinis, arba spalvų transformacijos, metodas (COCOA-1). Duomenų transformacijos tikslu taikomi trys plačiai naudojami spalvų DPM modeliai, nagrinėti ankstesnėje literatūroje. Tačiau šio tyrimo metu buvo analizuoti ir tirti rgb2lab bei pilkos skalės papildymo metodai. Egzistuoja 256 pilkos spalvos atspalviai, kurių šviesumas skiriasi nuo 0 (juoda) iki 1 (balta). Šie virsmo metodai dar vadinami monochromija. Kiti spalvų virsmo metodai, naudoti mokymo duomenims papildyti, yra kontrastas, rgb2lab, šviesumas ir rgb2gray.

2. Triukšmo įvedimo metodas (COCOA-2). Labiausiai naudojamas intensyvumo pakeitimo metodas yra triukšmo įvedimo metodas (NIM). Bet koks vaizdo signalo suprastėjimas, kurį dažnai sukelia išoriniai trikdžiai, vadinamas triukšmu vaizduose. Taigi, kai mokymo metu į vaizdų duomenis įvedamas triukšmas, tai labai paveikia mokymosi modelius, nes modelis tampa veiksmingesnis, o mokymasis geresnis [256]. Norint sukurti dirbtinius duomenų rinkinius, kiekvienam duomenų rinkinio vaizdui taikyti skirtingi triukšmo spektrai, pavyzdžiui, druskos ir pipirų triukšmas, Gauso triukšmas.
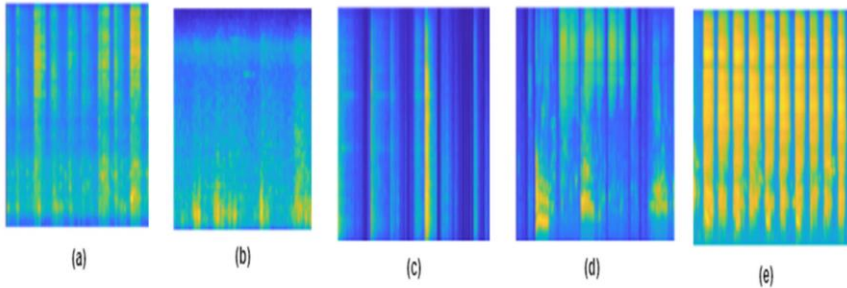
7.5 pav. Mel-spektrogramos vaizdų klasės: (a) teigiamas besimptomis; (b) teigiamas lengvas; (c) teigiamas vidutinis; (d) visiškas pasveikimas; (e) sveikas.
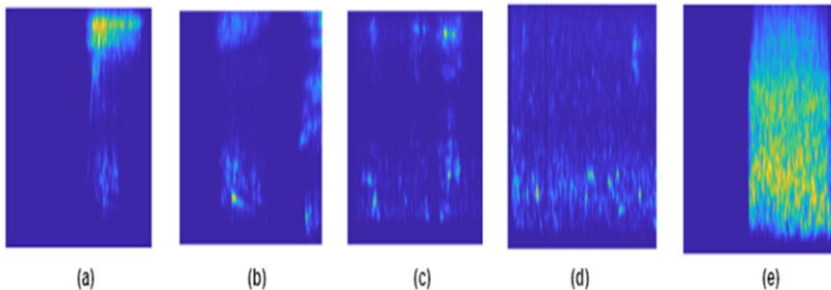
7.6 pav. GFCC vaizdų klasės: (a) teigiamas besimptomis; (b) teigiamas lengvas; (c) teigiamas vidutinis; (d) visiškas pasveikimas; (e) sveikas.

7.4.5. Veikimo metrika

Buvo įvertinti siūlomo DPM poveikio mokymosi modeliui rezultatai. Norint apskaičiuoti siūlomų modelių veiksmingumą, sukurta standartinė metrika, t. y. TP – teisingi teigiami, FP – klaidingi teigiami, TN – teisingi neigiami, FN – klaidingi neigiami. Naudojama ir ši optimizavimo metrika: ACC – tikslumas, SEN – jautrumas arba atkūrimas, SPEC – specifiškumas, PRE – preciziškumas, F1 balas, imtuvo darbinių charakteristikų kreivė / zona po kreive (ROC-AUC).

7.5. VYKDYMO IR ANALIZĖS ATLIKIMO REZULTATAI
7.5.1. Eksperimentiniai veido paralyžiaus diagnozavimo rezultatai

Eksperimentui naudotas hibridinis SqueezeNet/ECOC-SVM klasifikatorius. Taikytas tiek 1 kadro, tiek 2 kadrų mokymasis su siūlomu VDRRE metodu. Geriausi rezultatai gauti integravus 2 kadrų mokymąsi į VDRRE. Preciziškumas, atkūrimas, F1 balas ir tikslumas buvo lygūs 99,35 %, 99,74 %, 99,54 % ir 99,35 % (žr. 7.2 lentelę). Palyginti su eksperimentiniais 1 kadro mokymosi duomenimis, veikimas buvo prastesnis, o preciziškumas, atkūrimas, F1 balas ir tikslumas buvo lygūs 98,85 %, 99,28 %, 99,72 % ir 99,07 %. Vis dėlto, netaikant papildymo, gauti tokie preciziškumo, atkūrimo, F1 balo ir tikslumo rezultatai: 85,59 %, 81,06 %, 91,85 % ir 78,62 %. Kita vertus, tiek 1 kadro, tiek 2 kadrų mokymasis su VDRRE papildymu lėmė ryškų rezultatų pagerėjimą (žr. 7.7pav.). Nelinijinis matmenų mažinimo metodas, žinomas kaip t-SNE, leidžia vizualiai pristatyti dvimačius duomenis kaip dvimatį žemėlapį, pateikiamą 7.8 pav., a. Zonos po kreive (AUC) rezultatai pateikiami 7.8 pav., b.

7.2 Lentelė. Eksperimentiniai hibridinio klasifikatoriaus, skirto paralyžiui diagnozuoti, rezultatai

Metodai	Vidutiniai klasifikavimo rezultatai (95% CI)			
	Tikslumas (%)	Atkūrimas (%)	Preciziškumas (%)	F1 balas (%)
Be papildymo	78,62 (SD: 5,65)	91,85(SD: 5,41)	81,06(SD: 4,50)	85,59(SD: 3,78)
Atsitiktinio ištrynimo papildymas	92,91(SD: 1,12)	96,14(SD: 0,83)	93,96(SD: 1,87)	95,04(SD: 1,42)
2 kadrų mokymasis +VDRRE	**99,35(SD: 0,24)**	**99,74(SD: 0,17)**	**99,35(SD: 0,24)**	**99,54(SD: 0,16)**
1 kadro mokymasis + VDRRE	99,07(SD: 0,60)	99,72(SD:0,28)	98,85(SD: 1,15)	99,28(SD:0,55)
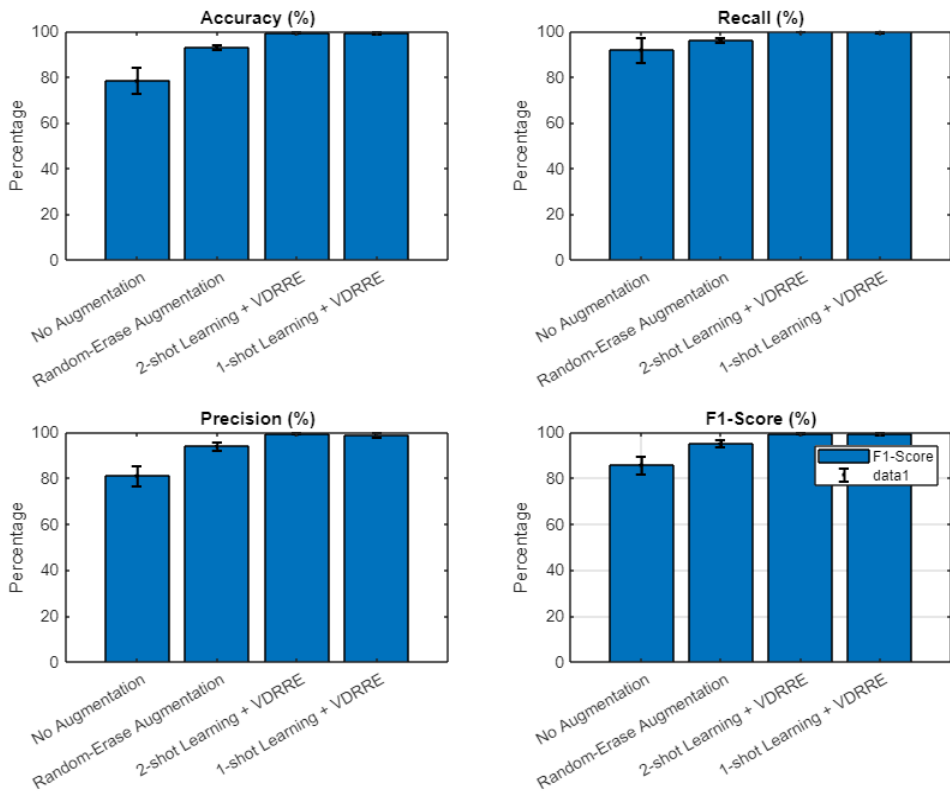
7.7 pav. Bendri dviejų pavyzdžių t-testo rezultatai atvejais be papildymo, 1 kadro ir 2 kadrų atvejais: (a) tikslumas; (b) atkūrimas; (c) preciziškumas ir (d) F1 balas
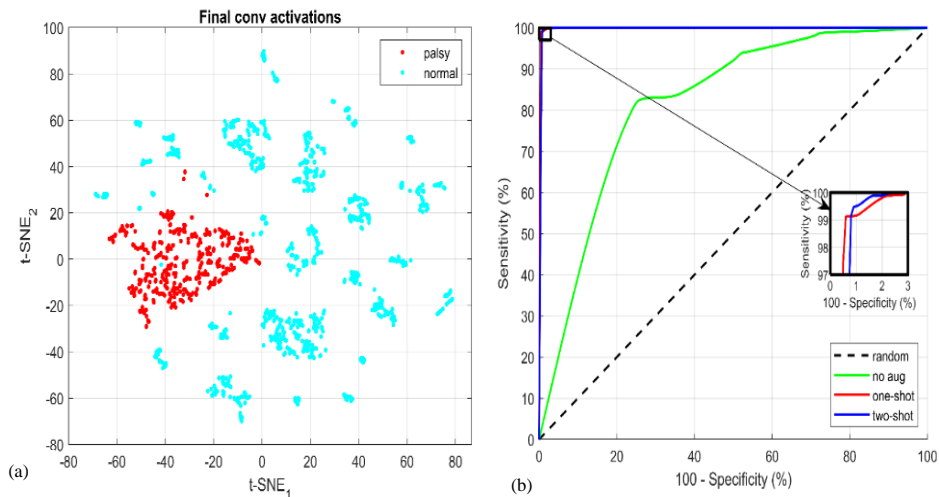
7.8 pav. (a) Duomenų vizualizacija, naudojant t-SNE; **(b)** ROC kreivė, rodanti AUC
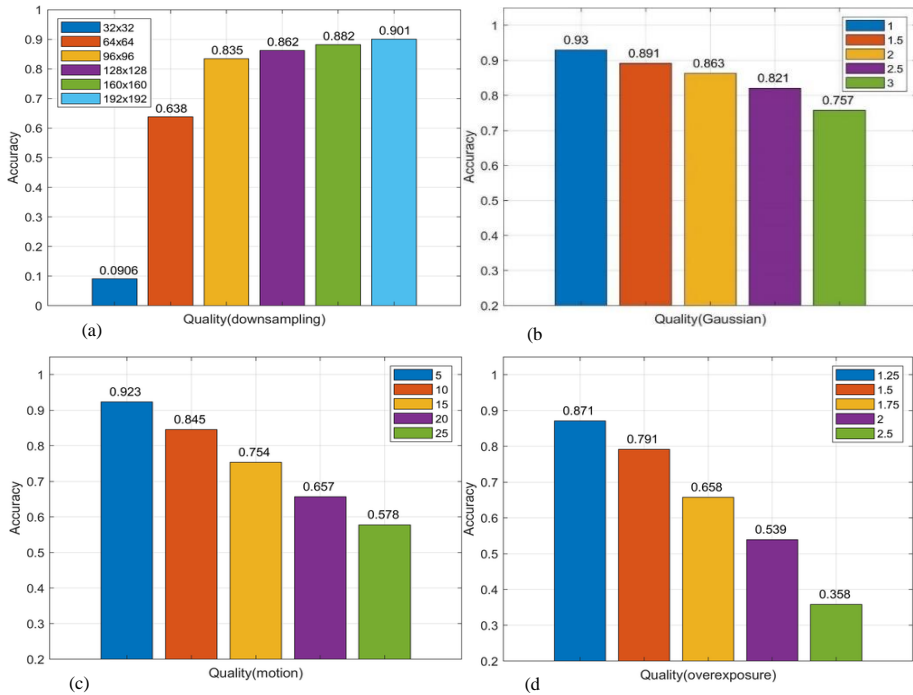
7.9 pav. Papildyti mokymo duomenų rinkinio veikimo rezultatai, kai k = 10: (a) skiriamosios gebos sumažinimas; **(b)** Gauso suliejimas; **(c)** judesio suliejimas; **(d)** perlaikymas

7.5.2. Eksperimentiniai maniokų ligų aptikimo rezultatai

Šiuo atveju buvo naudotas pradinio mokymo duomenų rinkinys su dirbtinai sukurtais vaizdais, naudojant histrogramų transformaciją, aprašytą 7.4.2 skyriuje. Naudotas nervinio tinklo modelis. Sujungus paryškintų vaizdų duomenų rinkinį su visais neapdorotais vaizdais, gauta 94350 vaizdų, naudotų baziniam tinkui performuoti. 80% vaizdų naudoti mokymui, o 20% – patvirtinimui. Performuoto tinklo modelis testuotas su neapdorotais duomenimis, aukštos ir žemos kokybės maniokų vaizdais. Dirbant su aukštesnės kokybės vaizdais, pasiektas 0,997 tikslumas. Jis pagerėjo 2%, tačiau pagerėjimas geriau matėsi prastesnės kokybės vaizduose. Sumažinus skiriamąją gebą, tikslumas padidėjo nuo 2,8% (192 ×192 px sumažinimas) iki 20,6% (64 ×64 px sumažinimas). Gauso suliejimo tikslumas išaugo nuo 2,8% (miglotumas prie 1°) iki 14,6% (miglotumas prie 3σ). Judėjimo suliejimo tikslumas išaugo nuo 4% (judėjimo suliejimas prie 5 px) iki 14,33% (judėjimo suliejimas prie 25 px). Perlaikymo tikslumas išaugo nuo 4% (2,5 kartus perlaikyta) iki 9,8% (25 kartus perlaikyta). 7.9 pav., a-d, pateikiami tinklo klasifikavimo rezultatai, kai vaizdai vertinti pagal Gauso suliejimą, judesio suliejimą ir perlaikymą.

Kai tinklas naudotas su išplėstiniu vaizdų duomenų bazės rinkiniu, sumažintų dirbtinių vaizdų rezultatai buvo geresni nei bazinio klasifikatoriaus, naudoto originaliems duomenų rinkinio vaizdams. Todėl siūlomo modelio o generalizacijos savybės buvo geresnės.

7.5.3. Eksperimentiniai odos melanomos diagnozavimo rezultatai

Pradinis duomenų rinkinys skirstomas į dvi dalis: dvinarė klasifikacija (melanoma ir ne melanoma) bei daugiaklasė klasifikacija (melanoma, netipiškų ir įprastų apgamų klasė). Kaip apibendrinta Tbl. 5.11 (žr. pagrindinėje disertacijos dalyje), dirbta su 120 pavyzdžių, kuriuos sudarė 24 melanomos ir 96 ne melanomos vaizdai (48 įprasti apgamai ir 48 netipiški apgamai). Testuota 80 vaizdų pavyzdžių, įskaitant 16 melanomos ir 64 ne melanomos vaizdus (sudaro 32 įprasti apgamai ir 32 netipiški apgamai). Mūsų siūlomam modeliui sukurti pasinaudota išankstinių mokymų tinklu. Siūlomas modelis pratestuotas su testuojamu duomenų rinkiniu, vidutiniu pradinio (bazinio) rinkinio veikimo rezultatu, AugDB-1 (taikytas CovSMOTE papildytų duomenų rinkinys) ir AugDB-2 (taikytas tradiciškai papildytų duomenų rinkinys), o rezultatai apibendrinami 7.3 lentelėje. 7.10 pav. lyginami veikimo rezultatai, naudojant didžiausius pasiektus rezultatus.
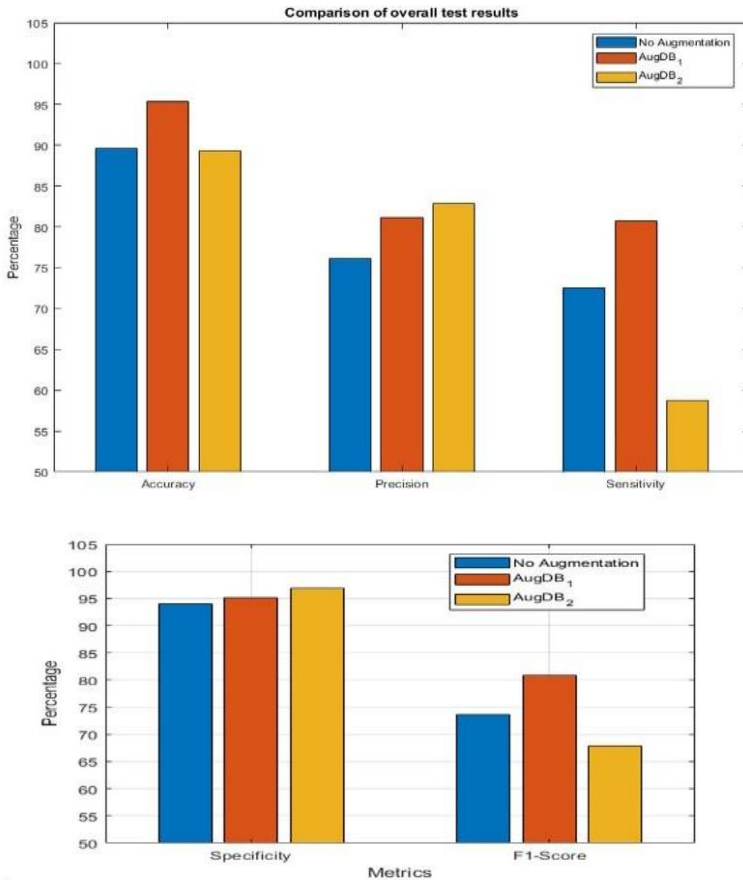
7.10 pav. Dvinarės klasifikacijos scenarijaus palyginimas

7.3 Lentelė. Odos melanomos klasifikavimo rezultatai (dvinarė klasifikacija*)*

Dvinarės klasės klasifikavimo rezultatai						
Duomenų rinkinys	**Aprašymas**	**Acc (%)**	**Pre (%)**	**Sen (%)**	**Spec (%)**	**F1 balas (%)**
Pradinis	PH^2	89,64	76,12	72,58	94,06	73,65
AugDB-1	CovSMOTE papildymas	95,31	81,13	80,77	95,1	80,84
AugDB-2	Tradicinis papildymas	89,26	82,87	58,75	96,88	67,83

Norint gauti patikimus rezultatus, kiti eksperimentai atlikti pradinėms trims PH2 duomenų rinkinio klasėms taikant daugiaklasį aptikimą. 7.11 pav. pateikiama eksperimentinių trijų klasių (melanomos, netipiškų apgamų ir įprastų apgamų) rezultatų santrauka.
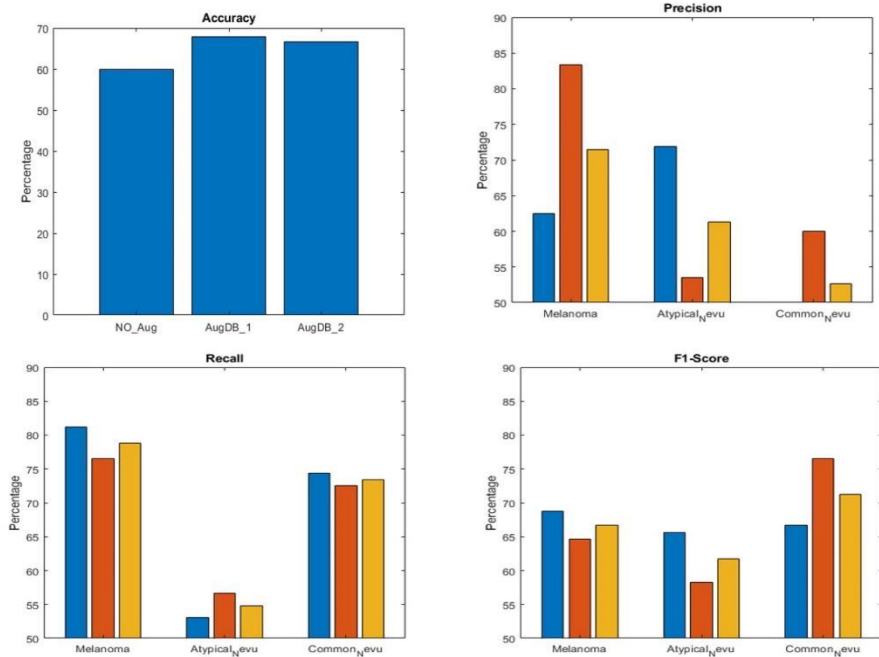
7.11 pav. Trijų duomenų rinkinių palyginimo rezultatai, taikant daugiaklasę klasifikaciją ir rodant tikslumą, jautrumą, preciziškumą, ir F1 balą

7.5.4. Eksperimentiniai Covid-19 diagnozavimo rezultatai (garso duomenų rinkinys)

Buvo atsižvelgta į detalią kiekvieno duomenų rinkinio analizę, naudojant siūlomą DeepShufNet, ir reikalavimus techninei įrangai. Tiek mokymui, tiek testavimui grupės dydis buvo sistemingai padidintas nuo 50 iki 200, laikant 200

priimtinu grupės dydžiu. Dėl didelio duomenų retumo „Coswara" duomenų rinkinyje eksperimentas pakartotas penkis kartus. Siūlomam DeepShufNet modeliui sukurti ir pratestuoti panaudoti jungtiniai savybių ištraukimo vaizdai iš visų „Coswara" duomenų rinkinių.

● Eksperimentinis rezultatas (sergantys COVID, palyginti su sveikais)

Pagal 7.4 lentelę, Mel-spektrogramos vaizdai COCOA-2 duomenų rinkiniuose pasižymėjo geriausiais rezultatais su DeepShufNet testuojamu rinkiniu. Teigiamas Covid klasifikavimo pavyzdys apskaičiuotas su vidutiniu tikslu ir standartiniu nukrypimu (SD): 85,1 (SD= 4.23), 70,85 (SD = 7,7) atkūrimo/jautrumo atveju, 59,64 (SD = 13,12) preciziškumo atveju, 88,25 (SD = 6,14) specifiškumo atveju ir 63,61 (SD = 6,7) F1-balo atveju. Pritaikius mūsų siūlomą modelį COCOA-3, tikslumas smarkiai padidėjo: vidutiniškai 87,82 (SD= 1,3), 69,49 (SD= 4,9) atkūrimo/jautrumo atveju, 64,82 (SD= 4,7) preciziškumo atveju, 91,75 (SD= 1,9) specifiškumo atveju ir FI balas kaip 66,9 (SD= 2,8). Todėl, lyginant rezultatus su kitais duomenų rinkiniais, galima teigti, kad bazinio rinkinio rezultatai yra mažiausiai veiksmingi.

GFCC vaizdų atveju papildyti duomenų rinkiniai pasižymėjo geresniais rezultatais nei baziniai, kalbant apie tikslumą: 83,1% (SD= 1,4) COCOA-3 atveju, 83,05% (SD= 0,9) COCOA-2 atveju, 76,4% (SD= 2,5) COCOA-1 atveju ir 74,9% (SD= 3,8) pradinių duomenų atveju 7.4 lentelė. Augantis vidutinis atkūrimas su DeepShufNet labiau intriguoja: 71,3% (SD= 2,2) COCOA-1, 48,7% (SD= 14,1) pradinių duomenų atveju, 46,7% (SD= 11,5) COCOA-2 atveju ir 38,8% (SD= 9.3) COCOA-3 atveju.
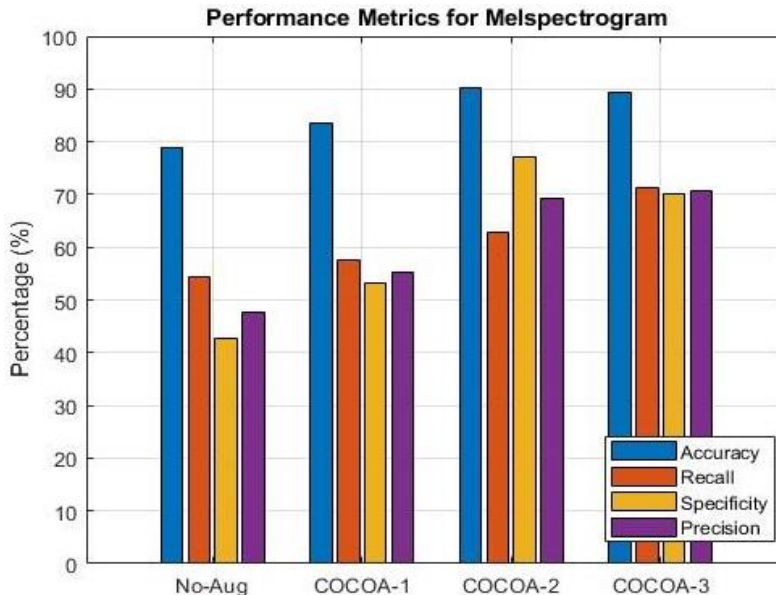
7.12 pav. Rezultatai, naudojant Mel spektrogramą (sveiki, palyginti su sergančiais COVID-19)

138
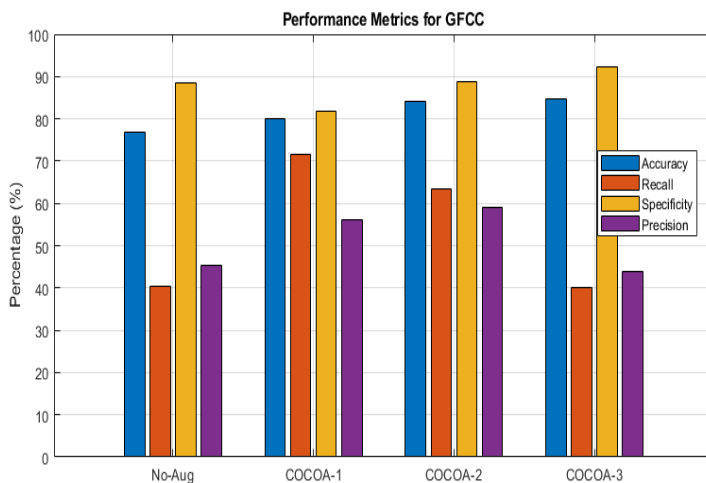
7.13 pav. Rezultatai, naudojant GFCC (sveiki, palyginti su sergančiais COVID-19)

7.4 Lentelė. Papildytų duomenų rinkinių klasifikavimo rezultatai (sergantys COVID palyginti su sveikais)

1 pvz.: visi teigiami COVID rezultatai palyginti su sveikais						
Savybių ištraukimas	**Duomenų rinkinys**	**Acc (%)**	**Rec (%)**	**Prec (%)**	**Spec (%)**	**F1 balas (%)**
Mel-spektrograma	Pradiniai duomenys	71,2±7,3	60,2±12,6	51,8±15,1	85,22±11,3	53,47±6,1
	COCOA-1	78,7±6,1	57,9±13,5	45,41±9,6	83,19±9,4	49,2±5,8
	COCOA-2	85,1±4,2	70,85±7,7	59,64±13,1	88,25±6,14	63,61±6,7
	COCOA-3	87,8±1,3	69,49±4,9	64,82±4,7	91,75±1,9	66,9±2,8
GFCC	Pradiniai duomenys	74,9±3,8	48,7±14,1	40,1±10,16	86,99±1,55	42,4±6,3
	COCOA-1	76,4±2,5	71,33±2,2	41,23±3,4	77,51±3,3	52,17±2,6
	COCOA-2	83,1±0,9	46,7±11,5	53,3±2,32	91,06±2,01	49,27±6,5
	COCOA-3	83,1±1,4	38,33±9,3	50,21±1,6	92,21±1,7	43,1±6,5

DeepShufNet su Mel-spektrogramos vaizdais analizė atskleidė, kad COCOA-2 eksperimentiniai rezultatai buvo didžiausi (pateikti 7.12 pav.): 90,1% tikslumas, 62,71% atkūrimas, 95,99% specifiškumas, 77,1%, ir 69,2% preciziškumas. Taikant duomenų papildymo strategiją, galima pagerinti veiklos rezultatus. Pavyzdžiui, dviejų COCOA-1 ir COCOA-2 rezultatai pasižymėjo geriausiu atkūrimu. Taip pat eksperimentiniai GFCC vaizdų rezultatai, pateikti 7.13 pav, rodo, kad taikant triukšmo papildymą COCOA-2 ir jungtinius duomenų rinkinius (COCOA-3), gautas didžiausias tikslumas – atitinkamai 84,1% ir 84,7%. Taigi, du geriausi atkūrimo rezultatai pasiekti su COCOA-1 ir COCOA-2, tai reiškia, kad duomenų papildymo metodas padeda pagerinti klasifikavimo rezultatus.

7.5.5. Visų siūlomų didinimo metodų analizė

Iš anksto išmokytame tinkle buvo išanalizuoti ir ištirti keturi šioje disertacijoje siūlomi duomenų papildymo metodai – VDRRE, CovSMOTE, fotometrinis (IQRM) ir triukšmo įpurškimo (NIM) metodai. Siekiant pateikti išsamų, lyginamąjį vertinimą, buvo lyginami trys veiksmingi iš anksto parengti tinklai: SqueezeNet, EfficientNet ir ResNet18 bei jų veikimas. Lyginant buvo siekiama nustatyti unikalias savybes, ir vėliau jas panaudoti norint tiksliai atpažinti ir klasifikuoti veido paralyžiaus ir odos melanomos ligas pagal siūlomą padidinimo metodą. Palyginimas buvo atliekamas dviejuose duomenų rinkiniuose, kad būtų galima visiškai suprasti kiekvieno papildymo metodo unikalumą, bei jų pranašumus ir trūkumus. Atlikus šį išsamų tyrimą, buvo atskleisti siūlomų DAM metodų ypatumai ir jų įtaka bendram efektyvumui kuriant geresnius giluminio mokymosi modelius.

Hiperparametrų nustatymai, naudojami iš anksto išmokytoms SqueezeNet, EfficientNet ir ResNet18 architektūroms optimizuoti, buvo sukurti taip, kad tiktų abiem duomenų rinkiniams (PH2 ir YFP). Visi nustatymai eksperimentams atlikti buvo kruopščiai sureguliuoti, siekiant užtikrinti, kad modelis būtų išmokytas teisingai identifikuoti ir suskirstyti į kategorijas odos melanomos ir veido paralyžiaus duomenų ligas. Hiperparametrų nustatymai ir jų reikšmės aprašyti 7.5 lentelėje.

7.5 Lentelė. Hiperparametrų nustatymai ir jų reikšmės.

Parametrai	Vertės		
	SqueezeNet	EfficientNet	ResNet18
Pradinis mokymosi tempas	0.00001	0.0001	0.003
Aktyvinimo funkcija	ReLU	ReLU	ReLU
Epochos	50	3	10
Partijos dydis	16	100	100
Optimizatorius	SGDM	ADAM	ADAM
L2 Reguliavimas	0.01	0.1	0.01
Impulsas	0.9	0.9	0.9
Svorio mokymosi koeficientas	20	10	10
Mokomi parametrai (milijonai)	1.2	5.3	11.6
Sluoksnių skaičius	63	290	71

Nustatant odos melanomos ligą, covSMOTE papildymo metodas pasiekė tiksliausius rezultatus iš visų trijų analizuotų klasifikatorių metodų. 7.6 lentelė parodo, kad covSMOTE padidinimas, pagrįstas ResNet18, pasiekė geriausius klasifikavimo rezultatus, palyginti su kitais šios kategorijos padidinimo metodų rezultatais 0,938 tikslumu, o VDRRE, PM ir NIM tikslumas atitinkamai buvo 0,925, 0,90 ir 0,912 bandymus atliekant tomis pačiomis sąlygomis. Iš anksto išmokytas tinklas SqueezeNet taip pat pasižymi gerais rezultatais, o geriausias rezultatas gautas

naudojant covSMOTE, kurio tikslumas yra 0,9218, NIM pasiekė 0,8522, PM – 0,8436, o VDRRE – 0,775 tikslumą atliekant eksperimentus. Siūlomas covSMOTE padidinimo metodas pasirodė labai efektyvus tiksliai klasifikuojant odos melanomos ligą, o tai rodo, kad nėra šališkumo jokiam mėginiui, kurio bendras atkūrimas, tikslumas, specifiškumas ir F1 balas yra atitinkamai 0,813, 0,867, 0,968 ir 0,839. Apskritai, siūlomas covSMOTE padidinimo modelis galėjo teisingai klasifikuoti daugiau paralyžiaus teigiamų klasės atvejų, kaip rodo tikslumo, atkūrimo ir preciziškumo pusiausvyra. Todėl galima teigti, kad trijų modelių (ypač ResNet18 ir SqeezeNet) išmokymas su covSMOTE pasižymėjo geru siūlomo modelio atsparumu ir generalizacijos gebėjimu tiksliai nustatyti ir klasifikuoti odos melanomą

Be to, veido paralyžiaus aptikimo atveju VDRRE pagrįstas padidinimas, palyginti su kitais siūlomais padidinimo metodais, rodo didelį skirtumą trijuose iš anksto išmokytuose tinkluose. 7.7 lentelėje apibendrinti eksperimentiniai rezultatai aiškiai parodo siūlomų DAM metodų testavimo efektyvumą nustatant veido paralyžių. Iš lentelėje esančių rezultatų galima daryti išvadą, kad VDRRE metodų bandymo tikslumas išlieka suderinamas su sklandžiu padidėjimu trijose iš anksto paruoštose CNN architektūrose, o didžiausias tikslumas siekia 0,9935 naudojant SqueezeNet, o EfficientNet ir ResNet architektūros bandymo rezultatai, naudojant VDRRE, yra atitinkamai 0,965 ir 0,929. Tačiau bandymo atlikimas naudojant covSMOTE rodo mažiausius rezultatus, palyginti su kitais siūlomais DAM metodais. Tačiau po penkių eksperimentinių bandymų ir patikslinimo, sumažinus mokymosi greitį iki 0,0003, pagerėjo bendra gilaus mokymosi modelio klasifikacija, todėl tikslumo rezultatas padidėjo +8,3%. Be to, fotometrinio ir NIM metodo taikymas rodo šiek tiek konservatyvų rezultatą ir sugebėjo sušvelninti perteklinį pritaikymą naudojant parametrų reguliavimą modelio mokymo metu. Apibendrinant galima teigti, kad eksperimentiniai rezultatai parodė labai gerus gebėjimus spręsti smulkias duomenų problemas, panaudojant konkrečios srities žinias, siekiant sėkmingai paveikti modelio mokymosi procesą ir taip įveikti mažų duomenų trūkumus realiose programose. Tačiau galima teigti, kad dėl duomenų ypatybių ir tikslinės srities unikalumo jokia viena papildymo technika negali būti laikoma geriausiu ar tobuliausiu metodu visais mažų duomenų atvejais.

7.5.6. Metodų apribojimas

Yra tam tikrų apribojimų, atsirandančių dėl labai mažų ir nesubalansuotų duomenų rinkinių, nes paprastai gilaus mokymosi modelių veiksmingumas ir geresni klasifikavimo rezultatai priklauso nuo didelių duomenų kiekio. CovSMOTE metodas pasižymėjo gerais rezultatais odos melanomos nustatymo atveju įvairiuose iš anksto paruoštuose modeliuose, tačiau gana žemais rezultatais veido paralyžiaus atveju. Tai gali būti siejama su duomenų klasių persidengimu, nes papildytuose vaizduose yra daugiau „triukšmo", todėl vaizdas atrodo nerealus. Be to, odos melanomos aptikimo scenarijuje iškilo modelio pertekliaus problema, nes bandymo rezultatai parodė tam tikrą šališkumą mažumos klasės atžvilgiu, ypač atliekant dvejetainę klasifikavimo užduotį. Be to, kelių klasių scenarijuje, nors modelis rodo optimalius rezultatus ir gana tiksliai klasifikuoja melanomą iš vienos klasės, jis gauna klasifikavimo klaidą diferencijuojant netipines (angl. *atypical nevu*) ir bendrines nevu (angl. *atypical nevu*)

klases, dėl kurių sumažėja atkūrimo (angl. *recall*) greitis ir specifiškumo rodiklis. Atliekant garso klasifikavimo užduotį, pagrindinis iššūkis yra aptikimo klaidos, kurios gali būti siejamos su prastu kai kurių sukurtų triukšmingų vaizdų apibendrinimu. Be to, buvo suvokta, kad kiekvienai garso klasei pritaikius bruožų išgavimo metodus, rodomų vaizdų galios atvaizdavimas yra labai panašus, todėl tai turi įtakos mokymosi klasifikatorių gebėjimui efektyviai apibendrinti duomenis.

7.6 Lentelė. Siūlomų odos melanomos klasifikacijos padidinimo metodų palyginimas.

Odos melanomos klasifikacija						
Duomenų papildymo metodai	Klasifikavimo modelis	Accuracy	Recall	Precision	Specificity	F1-Score
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	SqueezeNet	0,775	0,6875	0,4479	0,7969	0,5249
Kovariantas SMOTE (covSMOTE)		0,921	0,8077	0,8257	0,951	0,8084
Fotometrinė (spalvų transformacija)		0,8436	0,8021	0,6134	0,8542	0,6625
Triukšmo įpurškimo metodas (NIM)		0,8522	0,6667	0,6588	0,8984	0,6352
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	EfficientNet	0,862	0,50	0,7273	0,9531	0,5926
Kovariantas SMOTE (covSMOTE)		0,875	0,375	0,75	0,9688	0,500
Fotometrinė (spalvų transformacija)		0,838	1,000	0,5517	0,7969	0,7111
Triukšmo įpurškimo metodas (NIM)		0,85	0,625	0,625	0,9063	0,625
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	ResNet18	0,925	0,875	0,968	0,667	0,919
Kovariantas SMOTE (covSMOTE)		**0,938**	0,813	0,867	0,968	0,839
Fotometrinė (spalvų transformacija)		0,90	0,625	0,8333	0,9686	0,7142
Triukšmo įpurškimo metodas (NIM)		0,912	0,737	0,875	0,9672	0,80

7.7 Lentelė. Siūlomų veido paralyžiaus ligų klasifikavimo padidinimo metodų palyginimas.

Veido paralyžiaus aptikimas						
Duomenų papildymo metodai (DAM)	Klasifikavimo modelis	Accuracy	Recall	Precision	Specificity	F1-Score
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	SqueezeNet	**0,9935**	**0,9974**	**0,9935**	**NA**	**0,9954**
Kovariantas SMOTE (covSMOTE)		0,662	0,525	1,00	1,00	0,688
Fotometrinė (spalvų transformacija)		0,801	0,991	0,785	0,333	0,876
Triukšmo įpurškimo metodas (NIM)		0,791	0,8959	0,825	0,5333	0,8590
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	EfficientNet	0,965	0,9638	0,9861	0,9667	0,9748
Kovariantas SMOTE (covSMOTE)		0,675	0,7692	0,7727	0,4444	0,7710
Fotometrinė (spalvų transformacija)		0,720	0,9321	0,7410	0,20	0,8826
Triukšmo įpurškimo metodas (NIM)		0,875	0,8778	0,9417	0,8605	0,9087
Voronojaus dekompozicijos atsitiktinio regiono trynima (VDRRE)	ResNet18	0,929	0,914	0,9854	0,9667	0,9483
Kovariantas SMOTE (covSMOTE)		0,579	0,407	1,000	1,000	0,5788
Fotometrinė (spalvų transformacija)		0,624	0,4932	0,9561	0,9444	0,6507
Triukšmo įpurškimo metodas (NIM)		0,768	0,7919	0,8706	0,7111	0,8294

7.5.7. Eksperimentų santrauka

Šiame skyriuje pateikiami skirtingų duomenų papildymo modelių ir mokymosi klasifikatorių taikymo įvairiose klasifikavimo užduotyse išvados ir eksperimentiniai rezultatai. Naudojant spalvų histogramos padidinimo metodą, mokymosi modelis (MobileNetV2) sugebėjo gauti 99,7 % patobulintą aukštos kokybės vaizdų tikslumą. Treniruotės su Gauso triukšmo padidinimu pasižymėjo geriausiu bendru klasifikacijos rezultatu. Tolesnis našumo palyginimas su esamais metodais rodo, kad siūlomas modelis yra veiksmingas norint aptikti manioko ligą. Rezultatas demonstruoja statistiškai reikšmingą manioko lapų ligos aptikimo gebėjimo pagerėjimą, kurį rodo patobulintas MobileNetV2 neuroninis tinklas. Pagrindinis praktinių duomenų rinkimo aspektas yra modelio naudojimo paprastumas nustatant ir diagnozuojant manioko lapų ligas prastesnės kokybės vaizduose.

Veido paralyžiui aptikti įdiegus siūlomą covSMOTE modelį buvo sukurti nauji sintetiniai duomenų rinkiniai ir pagerėjo bendras bandymo duomenų rinkinio veikimas. Įdiegtas shuffleNet ir SVM aptikimo algoritmas pasirodė našesnis, pagerino tikslumą, jautrumą / atkūrimo greitį, specifiškumą ir mažesnį laiko sudėtingumą. Galiausiai, mūsų siūlomas modelis suteikė geresnių įžvalgų apie sumažintą klaidų dažnį ir apibendrinimo galimybes veido paralyžiaus aptikimo tikslumu 99,35 %, prisiminimo rodikliu 99,74 %, tikslumo koeficientu 99,35 %b ir Fi-balu 99,54 %.

Odos melanomos aptikimo tikslumas (92,18 %), jautrumas (80,77 %), specifiškumas (95,1 %) ir F1 balas (80,84 %), dvejetainio klasifikavimo scenarijaus eksperimentiniai rezultatai rodo, kad melanomos diagnozė gerokai pagerėjo. Kelių klasių klasifikavimo eksperimentiniai rezultatai parodė reikšmingą 89,2 % (jautrumas), 96,2 % (specifiškumas) netipinio *nevu* aptikimo, 65,4 % (jautrumo), 72,2 % (specifiškumas) ir 66 % (jautrumo) bendrojo *nevu* aptikimo efektyvumą. Kalbant apie odos melanomos aptikimą, siūloma klasifikavimo sistema pranoksta kai kuriuos ankstesnius metodus.

COVID-19 aptikti naudojant garso duomenų rinkinius buvo pritaikyti geometriniai ir fotometriniai DeepShufNet modelio padidinimo metodai. Naudojant Mel COCOA-2 patobulintus mokymo duomenų rinkinius, siūlomas modelis aptiko teigiamą COVID-19, kurio tikslumas, tikslumas, prisiminimas, specifiškumas ir f balas buvo atitinkamai 90,1 %, 77,1 %, 62,7 %, 95,98 % ir 69,1 %. Palyginus siūlomą metodą su kai kuriais naujausiais metodais, jis pasirodė našesnis.

Galiausiai būtina nustatyti, kad klasifikavimo užduotyse daugiau duomenų poreikio negalima per daug pabrėžti, tačiau pateikti mūsų siūlomų giluminio mokymosi modelių papildymo metodų rezultatai rodo daug žadančius rezultatus.

7.6. IŠVADOS IR REKOMENDACIJOS ATEITIES DARBAMS

7.6.1. Bendrosios išvados

Šioje disertacijoje buvo atliktas išsamus duomenų papildymo metodų taikymo tyrimas. DAM taikymas padėjo tobulinti mokymosi modelius, padidindamas vaizdų klasifikavimo užduočių apibendrinimą mažų duomenų analizei.

1. Dirbtinio intelekto metodų panaudojamumas, ypač nustatant medicinines ligas, ir toliau pritraukia vis daugiau dėmesio. Tačiau vis dar yra veiksnių, turinčių įtakos bendriems mažų duomenų analizės srities tyrimams ir naudojimui, ypač atliekant kompiuterio vizualizavimo užduotis. Vieni iš pagrindinių iššūkių yra didelė priklausomybė nuo didelių duomenų kiekių rinkinių; funkcijų atvaizdavimo problemos, kylančios dėl skirtingų vaizdo savybių skirtumų; duomenų retumo problema; atrankos paklaida; disbalanso klasės, tikimybė, kad modeliai bus pritaikyti per daug. Visa tai lemia mažesnius šansus, kad egzistuojantys sprendimai bus pritaikyti realiame gyvenime. Duomenų papildymo metodų taikymas labai padėjo pasiekti pageidaujamą mokymo duomenų dispersiją, o tai leido sukurti efektyvų mokymosi modelį ir taip pagerinti klasifikatorių veikimo rezultatus. Kai kurie esami duomenų papildymo metodų iššūkiai apima nerealių papildytų vaizdų problemą, kuri neturi reikšmingos įtakos klasifikavimo modelių veikimui.

2. Buvo sukurtas veiksmingas duomenų papildymo metodas, grindžiamas naudojamų duomenų išskirtinumu. Vėliau išnagrinėtas penkių klasifikavimo uždavinių gilaus mokymosi modelių veiksmingumas. Čia pateikiami apibendrinti rezultatai:

   i. Pasiūlytas naujas DPM metodas, pagrįstas stačiakampių Čebyševo funkcijų susiejimu su tikimybių pasiskirstymo funkcijomis (PDF), aptinkant maniokų ligas. Kaip dirbtinių mokymo duomenų rinkinių sudarymo strategija, mažinant vaizdų kokybę, panaudoti keturi papildymo būdai: judesio suliejimas, Gauso suliejimas, skiriamosios gebos mažinimas ir perlaikymas. Veiksmingai klasifikacijai taikytas ekonomiškas ir mažiau skaičiavimų reikalaujantis gilaus nervinio tinklo modelis (MobileNetV2).

   ii. Pristatytas VDRRE vaizdų papildymo metodas grindžiamas netaisyklingomis sritimis, sudarytomis taikant Vorronojaus mozaiką ir atsitiktinio ištrynimo papildymą, kuriant dirbtinius vaizdus veido paralyžiaus diagnozavimo tikslu. Be to, pasiūlytas mokymasis keliais kadrais vieno arba dviejų kadrų pagrindu, leidžiantis protingai klasifikuoti veido paralyžiaus ligas.

   iii. Sprendžiant duomenų trūkumo ir klasių disbalanso problemą odos melanomos diagnozavimo atveju, panaudotas kovariantinis dirbtinio mažumos įrašų kūrimo metodas (covSMOTE). Siūlomas metodas grįstas per didele duomenų atranka nelinijinėse mažesnių matmenų įterptinėse kopijose papildytų vaizdų sukūrimo tikslu. DPM naudojamas naujiems papildytiems vaizdams, skirtiems SqueezeNet gilaus mokymosi modeliui sukurti.

<ol type="i" start="22">
Covid-19 per gilaus kvėpavimo garsams diagnozuoti pasiūlytas triukšmo įvedimo ir spalvų transformacija grindžiamas papildymo metodas. Šios papildymo strategijos panaudotos kuriant dirbtinius mokymo duomenų rinkinius. Buvo taikomi du savybių ištraukimo metodai: Mel spektrograma ir GFCC.

3. Buvo pagerintas mokymosi modelių tikslumas, atkūrimas, preciziškumas ir specifiškumas mažų duomenų analizės srityje.

<ol type="i">
Maniokų ligų aptikimo atveju: pagrindinio tinklo, naudoto su prastos kokybės vaizdais, tikslumas pasiekė 97,7 %. Tačiau siūlomo modelio rezultatas padidėjo +2,0 %, panaudojus pradinius (aukštos kokybės) ir papildytus duomenų rinkinių (žemos kokybės) vaizdus su 99,7 % tikslumu. Taigi, rezultatai buvo geresni nei esamų moderniausių metodų.
Diagnozuojant veido paralyžių, didžiausias tikslumas pasiektas naudojant dviejų kadrų mokymąsi su siūlomu VDRRE metodu ir hibridiniu SqueezeNet/ECOC-SVM klasifikatoriumi. Mūsų siūlomas modelis pasiekė 99,34 % tikslumą. Mokymosi vienu kadru atveju tikslumas šiek tiek sumažėjo – nuo 99,34 % iki 99,07 %. Palyginti su moderniausiais metodais, mūsų siūlomas modelis smarkiai padidino tikslumą, preciziškumą ir atkūrimą.
Eksperimentiniai odos melanomos diagnozavimo užduoties rezultatai buvo įspūdingi dvinarės klasifikacijos klasės atveju. 92,18 % tikslumas, aptinkant melanomas, pastebėtas vykdant dvinario aptikimo užduotį, tačiau geriausias 89,2 % tikslumas pasiektas taikant daugiaklasės klasifikacijos scenarijų.
Diagnozuojant Covid-19, taikytas triukšmo įvedimo metodas. Naudojant Mel spektrogramą ir GFCC, išaugo klasifikacijos tikslumas. Tikslumu grindžiamas veiklos vertinimas rodo, kad mūsų siūlomas papildymas su DeepShufNet leido pasiekti geriausius eksperimentinius rezultatus, palyginti su moderniausiais metodais. Pasiektas 90,1 % tikslumas.

4. Tolesnis papildymo metodų veiksmingumo tyrimas parodė reikšmingus eksperimentinius rezultatus panaudojant konkrečios srities žinių aktualumą, siekiant pagerinti modelio mokymosi proceso apibendrinimą realiame gyvenime.

7.6.2. Rekomendacijos ateities darbams

Remiantis šios disertacijos išvadomis apie duomenų papildymo būdus ir jų taikymą įvairiose klasifikavimo užduotyse, siūlomos šios rekomendacijos būsimiems tyrimams:

1. Ištirti papildomus duomenų papildymo būdus, pavyzdžiui, pažangias geometrines transformacijas ir generatyvius priešingus tinklus (GAN), skirtus sintetiniams duomenims generuoti. Ištirti AI integravimą su kitomis naujomis technologijomis, pavyzdžiui, papildyta realybe duomenims papildyti.
2. Išplėsti siūlomus metodus į kitas sritis, kuriose duomenų trūkumas yra problema, pavyzdžiui, aplinkos stebėjimo, finansinio modeliavimo, emocijų atpažinimo ar socialinių mokslų.

3. Atlikti patikimą siūlomų duomenų papildymo metodų patvirtinimą įvairiuose duomenų rinkiniuose ir realaus pasaulio scenarijuose, atsižvelgiant į skirtingas duomenų charakteristikas bei aplinkos kintamuosius.
4. Atsižvelgiant į „triukšmingų duomenų" poveikį klasifikatoriaus veikimui, reiktų paieškoti / sukurti pažangesnius triukšmo mažinimo ir duomenų papildymo tvarkymo metodus, ypač atliekant garso ir vaizdo klasifikavimo užduotis.

REFERENCES

1. XU, Yongjun, LIU, Xin, CAO, Xin, HUANG, Changping, LIU, Enke, QIAN, Sen, LIU, Xingchen, WU, Yanjun, DONG, Fengliang and QIU, Cheng-Wei. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*. 20s21. Vol. 2, no. 4.

2. ZHANG, Huiyun, HUANG, Heming and HAN, Henry. MA-CapsNet-DA: Speech emotion recognition based on MA-CapsNet using data augmentation. *Expert Systems with Applications*. Online. 2024. Vol. 244, p. 122939. DOI https://doi.org/10.1016/j.eswa.2023.122939.

3. BEGUM, Ninja and HAZARIKA, Manuj Kumar. Artificial Intelligence in Agri-Food Systems—An Introduction. *Internet of Things and Analytics for Agriculture, Volume 3*. 2022. P. 45–63.

4. SIVAKUMAR, Jayanth, RAMAMURTHY, Karthik, RADHAKRISHNAN, Menaka and WON, Daehan. Synthetic sampling from small datasets: A modified mega-trend diffusion approach using k-nearest neighbors. *Knowledge-based systems*. 2022. Vol. 236, p. 107687.

5. CHEN, Hung-Yu, LI, Der-Chiang and LIN, Liang-Sian. Extending sample information for small data set prediction. In: *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, 2016. p. 710–714. ISBN 1467389854.

6. HE, Yan-Lin, WANG, Ping-Jiang, ZHANG, Ming-Qing, ZHU, Qun-Xiong and XU, Yuan. A novel and effective nonlinear interpolation virtual sample generation method for enhancing energy prediction and analysis on small data problem: A case study of Ethylene industry. *Energy*. 2018. Vol. 147, p. 418–427.

7. LINDSTROM, Martin. *Small data: The tiny clues that uncover huge trends*. St. Martin's Press, 2016. ISBN 1466892595.

8. CHANG, Che Jung, LI, Der Chiang, HUANG, Yi Hsiang and CHEN, Chien Chih. A novel gray forecasting model based on the box plot for small manufacturing data sets. *Applied Mathematics and Computation*. 2015. Vol. 265, p. 400–408. DOI 10.1016/j.amc.2015.05.006.

9. FARAWAY, Julian J and AUGUSTIN, Nicole H. When small data beats big data. *Statistics & Probability Letters*. 2018. Vol. 136, p. 142–145.

10. YUDELSON, Michael, FANCSALI, Steve, RITTER, Steve, BERMAN, Susan, NIXON, Tristan and JOSHI, Ambarish. Better data beats big data. In: *Educational data mining 2014*. Citeseer, 2014.

11. MOTAMED, Saman, ROGALLA, Patrik and KHALVATI, Farzad. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Informatics in Medicine Unlocked*. 2021. Vol. 27, p. 100779.

12. KOKOL, Peter, KOKOL, Marko and ZAGORANSKI, Sašo. Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress*. 2022. Vol. 105, no. 1, p. 00368504211029777.

13. GARCEA, Fabio, SERRA, Alessio, LAMBERTI, Fabrizio and MORRA, Lia. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*. 2022. P. 106391.

14. ALI-GOMBE, Adamu, ELYAN, Eyad and JAYNE, Chrisina. Multiple fake classes GAN for data augmentation in face image dataset. In: *2019 International joint conference on neural networks (IJCNN)*. IEEE, 2019. p. 1–8. ISBN 1728119855.

15. GIRIDHARA, Praveen Kumar Badimala, MISHRA, Chinmaya,

VENKATARAMANA, Reddy Kumar Modam, BUKHARI, Syed Saqib and DENGEL, Andreas. A Study of Various Text Augmentation Techniques for Relation Classification in Free Text. *ICPRAM*. 2019. Vol. 3, p. 5.

16. HUSSAIN, Zeshan, GIMENEZ, Francisco, YI, Darvin and RUBIN, Daniel. Differential data augmentation techniques for medical imaging classification tasks. In: *AMIA annual symposium proceedings*. American Medical Informatics Association, 2017. p. 979.

17. RUSSAKOVSKY, Olga, DENG, Jia, SU, Hao, KRAUSE, Jonathan, SATHEESH, Sanjeev, MA, Sean, HUANG, Zhiheng, KARPATHY, Andrej, KHOSLA, Aditya and BERNSTEIN, Michael. Imagenet large-scale visual recognition challenge. *International journal of computer vision*. 2015. Vol. 115, p. 211–252.

18. PEREZ, Luis and WANG, Jason. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*. 2017.

19. ABBAS, Amreen, JAIN, Sweta, GOUR, Mahesh and VANKUDOTHU, Swetha. Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Computers and Electronics in Agriculture*. 2021. Vol. 187, p. 106279.

20. DORNAIKA, Fadi, BI, Jingjun and ZHANG, Chongsheng. A unified deep semi-supervised graph learning scheme based on nodes re-weighting and manifold regularization. *Neural Networks*. 2023. Vol. 158, p. 188–196.

21. LEE, Ji Won, NAM, Do-Won, YOO, Won-Young, KIM, Yoonhyung, JEONG, Minki and KIM, Changick. Soccer object motion recognition based on 3D convolutional neural networks. In: *FedCSIS (Communication Papers)*. 2018. p. 129–134.

22. OBERMEYER, Ziad and EMANUEL, Ezekiel J. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*. 2016. Vol. 375, no. 13, p. 1216.

23. FUJISAWA, Yasuhiro, INOUE, Sae and NAKAMURA, Yoshiyuki. The possibility of deep learning-based, computer-aided skin tumour classifiers. *Frontiers in medicine*. 2019. Vol. 6, p. 191.

24. MIRANDA, Eka, ARYUNI, Mediana and IRWANSYAH, E. A survey of medical image classification techniques. In: *2016 international conference on information management and technology (ICIMTech)*. IEEE, 2016. p. 56–61. ISBN 1509033521.

25. LU, Rui, DUAN, Zhiyao and ZHANG, Changshui. Metric learning-based data augmentation for environmental sound classification. In: *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2017. p. 1–5. ISBN 1538616327.

26. TRUONG, Thanh-Nghia, DAM, Vu-Duy and LE, Thanh-Sach. Medical image sequence normalization and augmentation: improve liver tumour segmentation from the small dataset. In: *2018 3rd International Conference on Control, Robotics and Cybernetics (CRC)*. IEEE, 2018. p. 1–5. ISBN 1538677385.

27. KOLÁŘ, Martin, HRADIŠ, Michal and ZEMČÍK, Pavel. Deep learning on small datasets using online image search. In: *Proceedings of the 32nd Spring Conference on Computer Graphics*. 2016. p. 87–93.

28. LI, Wei, CHEN, Chen, ZHANG, Mengmeng, LI, Hengchao and DU, Qian. Data augmentation for hyperspectral image classification with deep CNN. *IEEE Geoscience and Remote Sensing Letters*. 2018. Vol. 16, no. 4, p. 593–597.

29. TAKALKAR, Madhumita A and XU, Min. Image-based facial micro-expression recognition using deep learning on small datasets. In: *2017 international conference on digital image computing: techniques and applications (DICTA)*. IEEE, 2017. p. 1–7. ISBN 1538628392.

30. ZHANG, Baohua, FAN, Shuxiang, LI, Jiangbo, HUANG, Wenqian, ZHAO, Chunjiang, QIAN, Man and ZHENG, Ling. Detection of early rottenness on apples by using hyperspectral imaging combined with spectral analysis and image processing. *Food Analytical Methods*. 2015. Vol. 8, p. 2075–2086.

31. KANG, Jaeyong and GWAK, Jeonghwan. Ensemble of multi-task deep convolutional neural networks using transfer learning for fruit freshness classification. *Multimedia Tools and Applications*. 2022. Vol. 81, no. 16, p. 22355–22377.

32. HAN, Yifei, LIU, Zhaojing, KHOSHELHAM, Kourosh and BAI, Shahla Hosseini. Quality estimation of nuts using deep learning classification of hyperspectral imagery. *Computers and Electronics in Agriculture*. 2021. Vol. 180, p. 105868.

33. MONTEIRO, Miguel, FONSECA, Ana Catarina, FREITAS, Ana Teresa, E MELO, Teresa Pinho, FRANCISCO, Alexandre P, FERRO, Jose M and OLIVEIRA, Arlindo L. Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEE/ACM transactions on computational biology and bioinformatics*. 2018. Vol. 15, no. 6, p. 1953–1959.

34. NIE, Dong, SHANK, Elizabeth A and JOJIC, Vladimir. A deep framework for bacterial image segmentation and classification. In: *Proceedings of the 6th ACM conference on bioinformatics, computational biology and health informatics*. 2015. p. 306–314.

35. LI, Der-Chiang, LIN, Wu-Kuo, LIN, Liang-Sian, CHEN, Chien-Chih and HUANG, Wen-Ting. The attribute-trend-similarity method to improve learning performance for small datasets. *International Journal of Production Research*. 2017. Vol. 55, no. 7, p. 1898–1913.

36. LIU, Bo, WEI, Ying, ZHANG, Yu and YANG, Qiang. Deep Neural Networks for High Dimension, Low Sample Size Data. In: *IJCAI*. 2017. p. 2287–2293.

37. NAYAK, Janmenjoy, VAKULA, Kanithi, DINESH, Paidi, NAIK, Bighnaraj and PELUSI, Danilo. Intelligent food processing: Journey from artificial neural network to deep learning. *Computer Science Review*. 2020. Vol. 38, p. 100297.

38. ZHANG, Jianpeng, XIE, Yutong, XIA, Yong and SHEN, Chunhua. Attention residual learning for skin lesion classification. *IEEE transactions on medical imaging*. 2019. Vol. 38, no. 9, p. 2092–2103.

39. KITCHIN, Rob and LAURIAULT, Tracey P. Small data in the era of big data. *GeoJournal*. 2015. Vol. 80, p. 463–475.

40. TSAI, Tung-I and LI, Der-Chiang. Utilize bootstrap in small data set learning for pilot run modelling of manufacturing systems. *Expert Systems with Applications*. 2008. Vol. 35, no. 3, p. 1293–1300.

41. LI, Der-Chiang and WEN, I-Hsiang. A genetic algorithm-based virtual sample generation technique to improve small data set learning. *Neurocomputing*. 2014. Vol. 143, p. 222–230.

42. ZHOU, Lina, PAN, Shimei, WANG, Jianwu and VASILAKOS, Athanasios V. Machine learning on big data: Opportunities and challenges. *Neurocomputing*. 2017. Vol. 237, p. 350–361.

43. RICHMOND, Brian. Introduction to Data Analysis Handbook. *Academy for Educational Development*. 2006.

44. WASIKOWSKI, Michael. *Combating the class imbalance problem in small sample data sets*. 2009. University of Kansas. ISBN 1109228481.

45. LATEH, Masitah Abdul, MUDA, Azah Kamilah, YUSOF, Zeratul Izzah Mohd, MUDA, Noor Azilah and AZMI, Mohd Sanusi. Handling a small dataset problem in prediction model by employing artificial data generation approach: A review. In: *Journal of Physics: Conference Series*. IOP Publishing, 2017. p. 012016. ISBN 1742-

6596.

46. MISHRA, Sourav, YAMASAKI, Toshihiko and IMAIZUMI, Hideaki. Improving image classifiers for small datasets by learning rate adaptations. In: *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 2019. p. 1–6. ISBN 4901122185.

47. LENG, Biao, YU, Kai and JINGYAN, Q I N. Data augmentation for unbalanced face recognition training sets. *Neurocomputing*. 2017. Vol. 235, p. 10–14.

48. BARZ, Bjorn and DENZLER, Joachim. Deep learning on small datasets without pre-training using cosine loss. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2020. p. 1371–1380.

49. WU, Yujie, DENG, Lei, LI, Guoqi, ZHU, Jun, XIE, Yuan and SHI, Luping. Direct training for spiking neural networks: Faster, larger, better. In: *Proceedings of the AAAI conference on artificial intelligence*. 2019. p. 1311–1318. ISBN 2374-3468.

50. CHEN, Hui-Ling, WANG, Gang, MA, Chao, CAI, Zhen-Nao, LIU, Wen-Bin and WANG, Su-Jing. An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. *Neurocomputing*. 2016. Vol. 184, p. 131–144.

51. KIM, Hazel H, WOO, Daecheol, OH, Seong Joon, CHA, Jeong-Won and HAN, Yo-Sub. Alp: Data augmentation using lexicalized pcfgs for few-shot text classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022. p. 10894–10902. ISBN 2374-3468.

52. VAN DYK, David A and MENG, Xiao-Li. The art of data augmentation. *Journal of Computational and Graphical Statistics*. 2001. Vol. 10, no. 1, p. 1–50.

53. HERNÁNDEZ-GARCÍA, Alex and KÖNIG, Peter. Data augmentation instead of explicit regularization. *arXiv preprint arXiv:1806.03852*. 2018.

54. HAN, Dongmei, LIU, Qigang and FAN, Weiguo. A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications*. 2018. Vol. 95, p. 43–56.

55. GÓMEZ-RÍOS, Anabel, TABIK, Siham, LUENGO, Julián, SHIHAVUDDIN, A S M, KRAWCZYK, Bartosz and HERRERA, Francisco. Towards highly accurate coral texture image classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications*. 2019. Vol. 118, p. 315–328.

56. SHORTEN, Connor, KHOSHGOFTAAR, Taghi M and FURHT, Borko. Text data augmentation for deep learning. *Journal of Big Data*. 2021. Vol. 8, p. 1–34.

57. FENG, Steven Y, GANGAL, Varun, WEI, Jason, CHANDAR, Sarath, VOSOUGHI, Soroush, MITAMURA, Teruko and HOVY, Eduard. A survey of data augmentation approaches for NLP. *arXiv preprint arXiv:2105.03075*. 2021.

58. PARK, Daniel S, CHAN, William, ZHANG, Yu, CHIU, Chung-Cheng, ZOPH, Barret, CUBUK, Ekin D and LE, Quoc V. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*. 2019.

59. KNOEDLER, Leonard, BAECHER, Helena, KAUKE-NAVARRO, Martin, PRANTL, Lukas, MACHENS, Hans-Günther, SCHEUERMANN, Philipp, PALM, Christoph, BAUMANN, Raphael, KEHRER, Andreas and PANAYI, Adriana C. Towards a reliable and rapid automated grading system in facial palsy patients: facial palsy surgery meets computer science. *Journal of Clinical Medicine*. 2022. Vol. 11, no. 17, p. 4998.

60. ABAYOMI-ALLI, Olusola Oluwakemi, DAMAŠEVIČIUS, Robertas, MISRA, Sanjay and MASKELIŪNAS, Rytis. Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. *Expert Systems*. 2021. Vol. 38, no. 7, p. e12746.

61. SHORTEN, Connor and KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. *Journal of big data*. 2019. Vol. 6, no. 1, p. 1–48.

62. IOFFE, Sergey and SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*. pmlr, 2015. p. 448–456.

63. SRIVASTAVA, Nitish, HINTON, Geoffrey, KRIZHEVSKY, Alex, SUTSKEVER, Ilya and SALAKHUTDINOV, Ruslan. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*. 2014. Vol. 15, no. 1, p. 1929–1958.

64. SIMONYAN, Karen and ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.

65. KRIZHEVSKY, Alex, SUTSKEVER, Ilya and HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012. Vol. 25.

66. ZHONG, Zhun, ZHENG, Liang, KANG, Guoliang, LI, Shaozi and YANG, Yi. Random erasing data augmentation. In: *Proceedings of the AAAI conference on artificial intelligence*. 2020. p. 13001–13008. ISBN 2374-3468.

67. YANG, Jiawen, XIE, Fengying, FAN, Haidi, JIANG, Zhiguo and LIU, Jie. Classification for dermoscopy images using convolutional neural networks based on region average pooling. *IEEE Access*. 2018. Vol. 6, p. 65130–65138.

68. WONG, Sebastien C, GATT, Adam, STAMATESCU, Victor and MCDONNELL, Mark D. Understanding data augmentation for classification: when to warp? In: *2016 international conference on digital image computing: techniques and applications (DICTA)*. IEEE, 2016. p. 1–6. ISBN 150902896X.

69. LIU, Jun S and WU, Ying Nian. Parameter expansion for data augmentation. *Journal of the American Statistical Association*. 1999. Vol. 94, no. 448, p. 1264–1274.

70. MURDOCK, Calvin, LI, Zhen, ZHOU, Howard and DUERIG, Tom. Blockout: Dynamic model selection for hierarchical deep networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 2583–2591.

71. DEVRIES, Terrance and TAYLOR, Graham W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*. 2017.

72. CARE, Algo, RAMPONI, Federico A and CAMPI, Marco C. A new classification algorithm with guaranteed sensitivity and specificity for medical applications. *IEEE Control Systems Letters*. 2018. Vol. 2, no. 3, p. 393–398.

73. JANG, Junwoo, HAN, Jungwoo and KIM, Jinwhan. K-mixup: Data augmentation for offline reinforcement learning using mixup in a Koopman invariant subspace. *Expert Systems with Applications*. 2023. Vol. 225, p. 120136.

74. ZHANG, Tao, LIN, Liqin and XUE, Zaifa. A voice feature extraction method based on fractional attribute topology for Parkinson's disease detection. *Expert Systems with Applications*. 2023. Vol. 219, p. 119650. DOI 10.1016/j.eswa.2023.119650.

75. CHEN, Zihan, QIAN, Yaojia, WANG, Yuxi and FANG, Yinfeng. Deep Convolutional Generative Adversarial Network-Based EMG Data Enhancement for Hand Motion Classification. *Frontiers in Bioengineering and Biotechnology*. 2022. Vol. 10, p. 909653.

76. ZHAO, Yi, TAKAKI, Shinji, LUONG, Hieu-Thi, YAMAGISHI, Junichi, SAITO, Daisuke and MINEMATSU, Nobuaki. Wasserstein GAN and waveform loss-based acoustic model training for multi-speaker text-to-speech synthesis systems using a WaveNet vocoder. *IEEE Access*. 2018. Vol. 6, p. 60478–60488.

77. PEREIRA, Sérgio, PINTO, Adriano, ALVES, Victor and SILVA, Carlos A. Brain tumour segmentation using convolutional neural networks in MRI images. *IEEE*

transactions on medical imaging. 2016. Vol. 35, no. 5, p. 1240–1251.

78. WANG, Guan, SUN, Yu and WANG, Jianxin. Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*. 2017. Vol. 2017.

79. GONG, Liang, DU, Xiaofeng, ZHU, Kai, LIN, Chenghui, LIN, Ke, WANG, Tao, LOU, Qiaojun, YUAN, Zheng, HUANG, Guoqiang and LIU, Chengliang. Pixel level segmentation of early-stage in-bag rice root for its architecture analysis. *Computers and Electronics in Agriculture*. 2021. Vol. 186, p. 106197. DOI 10.1016/j.compag.2021.106197.

80. LU, Yu, FU, Xianghua, LI, Xiaoqing and QI, Yingjian. Cardiac Chamber Segmentation Using Deep Learning on Magnetic Resonance Images from Patients before and after Atrial Septal Occlusion Surgery. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. IEEE, 2020. p. 1211–1216. ISBN 9781728119908. DOI 10.1109/EMBC44109.2020.9175618.

81. KOMPANEK, Matej, TAMAJKA, Martin and BENESOVA, Wanda. Volumetric Data Augmentation as an Effective Tool in MRI Classification Using 3D Convolutional Neural Network. In: *International Conference on Systems, Signals, and Image Processing*. IEEE, 2019. p. 115–119. ISBN 9781728132273. DOI 10.1109/IWSSIP.2019.8787315.

82. XIN, Zhimeng, LU, Tongwei, LI, Yuzhou and YOU, Xinge. MultiCut-MultiMix: a two-level data augmentation method for detecting small and densely distributed objects in large-size images. *Visual Computer*. 2023. P. 1–15. DOI 10.1007/s00371-023-02920-z.

83. GOODFELLOW, Ian. NIPS 2016 Tutorial: Generative Adversarial Networks. *arXiv preprint arXiv:1701.00160*. Online. 2016. Available from: http://arxiv.org/abs/1701.00160

84. KUO, Nicholas I.Hsien, GARCIA, Federico, SÖNNERBORG, Anders, BÖHM, Michael, KAISER, Rolf, ZAZZI, Maurizio, POLIZZOTTO, Mark, JORM, Louisa and BARBIERI, Sebastiano. Generating synthetic clinical data that capture class imbalanced distributions with generative adversarial networks: Example using antiretroviral therapy for HIV. *Journal of Biomedical Informatics*. 2023. Vol. 144, p. 104436. DOI 10.1016/j.jbi.2023.104436.

85. ZHANG, Xiaofeng, WANG, Zhangyang, LIU, Dong, LIN, Qifeng and LING, Qing. Deep Adversarial Data Augmentation for Extremely Low Data Regimes. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021. Vol. 31, no. 1, p. 15–28. DOI 10.1109/TCSVT.2020.2967419. D.

86. ESKIMEZ, Sefik Emre, DIMITRIADIS, Dimitrios, GMYR, Robert and KUMANATI, Kenichi. GAN-based data generation for speech emotion recognition. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. 2020. p. 3446–3450. DOI 10.21437/Interspeech.2020-2898.

87. Zhu, X., Liu, Y., Qin, Z., arXiv:1711.00648, J. L. preprint, & 2017, (2017). Data augmentation in emotion classification using generative adversarial networks. Arxiv.Org. https://arxiv.org/abs/1711.00648

88. ABAYOMI-ALLI, Olusola Oluwakemi, DAMAŠEVIČIUS, Robertas, WIECZOREK, Michał and WOŹNIAK, Marcin. Data augmentation using principal component resampling for image recognition by deep learning. In: *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020,*

Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19. Springer, 2020. p. 39–48. ISBN 3030615332.

89. ABAYOMI-ALLI, Olusola Oluwakemi, DAMAŠEVIČIUS, Robertas, MASKELIŪNAS, Rytis and MISRA, Sanjay. Few-shot learning with a novel Voronoi tessellation-based image augmentation method for facial palsy detection. *Electronics*. 2021. Vol. 10, no. 8, p. 978.

90. GILDEN, Donald H. Bell's palsy. *New England Journal of Medicine*. 2004. Vol. 351, no. 13, p. 1323–1331.

91. NELLIS, Jason C, ISHII, Masaru, BYRNE, Patrick J, BOAHENE, Kofi D O, DEY, Jacob K and ISHII, Lisa E. Association among facial paralysis, depression, and quality of life in facial plastic surgery patients. *JAMA facial plastic surgery*. 2017. Vol. 19, no. 3, p. 190–196.

92. LOU, Jianwen, YU, Hui and WANG, Fei-Yue. A review on automated facial nerve function assessment from visual face capture. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2019. Vol. 28, no. 2, p. 488–497.

93. KIHARA, Yuta, DUAN, Guifang, NISHIDA, Takeshi, MATSUSHIRO, Naoki and CHEN, Yen-Wei. A dynamic facial expression database for quantitative analysis of facial paralysis. In: *2011 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*. IEEE, 2011. p. 949–952. ISBN 8988678559.

94. BANKS, Caroline A, BHAMA, Prabhat K, PARK, Jong, HADLOCK, Charles R and HADLOCK, Tessa A. Clinician-graded electronic facial paralysis assessment: the eFACE. *Plastic and reconstructive surgery*. 2015. Vol. 136, no. 2, p. 223e-230e.

95. HE, Shu, SORAGHAN, John J, O'REILLY, Brian F and XING, Dongshan. Quantitative analysis of facial paralysis using local binary patterns in biomedical videos. *IEEE Transactions on Biomedical Engineering*. 2009. Vol. 56, no. 7, p. 1864–1870.

96. WANG, Li, LI, Rui-Feng, WANG, Ke and CHEN, Jian. Feature representation for facial expression recognition based on FACS and LBP. *International Journal of Automation and Computing*. 2014. Vol. 11, p. 459–468.

97. NGO, Truc Hung, SEO, Masataka, MATSUSHIRO, Naoki, XIONG, Wei and CHEN, Yen-Wei. Quantitative analysis of facial paralysis based on limited-orientation modified circular Gabor filters. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016. p. 349–354. ISBN 1509048472.

98. JIANG, Wei, ZHANG, Kai, WANG, Nan and YU, Miao. MeshCut data augmentation for deep learning in computer vision. *PLoS One*. 2020. Vol. 15, no. 12, p. e0243613.

99. HSU, Gee-Sern Jison, KANG, Jiunn-Horng and HUANG, Wen-Fong. Deep hierarchical network with line segment learning for quantitative analysis of facial palsy. *IEEE Access*. 2018. Vol. 7, p. 4833–4842.

100. REDMON, Joseph and FARHADI, Ali. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. 2018.

101. GUO, Demi, KIM, Yoon and RUSH, Alexander M. Sequence-level mixed sample data augmentation. *arXiv preprint arXiv:2011.09039*. 2020.

102. STOREY, Gary, JIANG, Richard, KEOGH, Shelagh, BOURIDANE, Ahmed and LI, Chang-Tsun. 3DPalsyNet: A facial palsy grading and motion recognition framework using fully 3D convolutional neural networks. *IEEE Access*. 2019. Vol. 7, p. 121655–121664.

103. GU, Ke, ZHANG, Yonghui and QIAO, Junfei. Ensemble meta-learning for few-shot soot density recognition. *IEEE Transactions on Industrial Informatics*. 2020. Vol. 17, no. 3, p. 2261–2270.

104. TEN HARKEL, Timen C, DE JONG, Guido, MARRES, Henri A M, INGELS, Koen J A O, SPEKSNIJDER, Caroline M and MAAL, Thomas J J. Automatic grading of patients with a unilateral facial paralysis based on the Sunnybrook Facial Grading System-A deep learning study based on a convolutional neural network. *American Journal of Otolaryngology*. 2023. Vol. 44, no. 3, p. 103810.

105. GABER, Amira, TAHER, Mona F, WAHED, Manal Abdel, SHALABY, Nevin Mohieldin and GABER, Sarah. Classification of facial paralysis based on machine learning techniques. *BioMedical Engineering OnLine*. 2022. Vol. 21, no. 1, p. 1–20.

106. PHIENPHANICH, Phongphan, LERTHIRUNVIBUL, Nichapa, CHARNNARONG, Ekabhat, MUNTHULI, Adirek, TANTIBUNDHIT, Charturong and SUWANWELA, Nijasri C. Generalizing a Small Facial Image Dataset Using Facial Generative Adversarial Networks for Stroke's Facial Weakness Screening. *IEEE Access*. 2023.

107. RUSIA, Mayank Kumar and SINGH, Dushyant Kumar. An efficient CNN approach for facial expression recognition with some measures of overfitting. *International journal of information technology*. 2021. Vol. 13, p. 2419–2430.

108. MWEBAZE, Ernest, GEBRU, Timnit, FROME, Andrea, NSUMBA, Solomon and TUSUBIRA, Jeremy. iCassava 2019 fine-grained visual categorization challenge. *arXiv preprint arXiv:1908.02900*. 2019.

109. ALENE, A., KHATAZA, R., CHIBWANA, C, NTAWURUHUNGA, Pheneas and MOYO, C. Economic impacts of cassava research and extension in Malawi and Zambia. *Journal of Development and Agricultural Economics*. 2013.

110. FELEKE, Shiferaw, MANYONG, Victor, ABDOULAYE, Tahirou and ALENE, Arega D. Assessing the impacts of cassava technology on poverty reduction in Africa. *Studies in Agricultural Economics*. 2016. Vol. 118, no. 2, p. 101–111.

111. CALZADILLA, Alvaro, ZHU, Tingju, REHDANZ, Katrin, TOL, Richard S J and RINGLER, Claudia. Economywide impacts of climate change on agriculture in Sub-Saharan Africa. *Ecological Economics*. 2013. Vol. 93, p. 150–165.

112. PEI, Y L, SHI, T, LI, C P, LIU, X B, CAI, J M and HUANG, G X. Distribution and pathogen identification of cassava brown leaf spot in China. *Genetics and Molecular Research*. 2014. Vol. 13, no. 2, p. 3461–3473.

113. NAZKI, Haseeb, YOON, Sook, FUENTES, Alvaro and PARK, Dong Sun. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Computers and Electronics in Agriculture*. 2020. Vol. 168, p. 105117.

114. KOLAWOLE, Peter O, AGBETOYE, Leo and OGUNLOWO, Simeon A. Sustaining world food security with improved cassava processing technology: The Nigeria experience. *Sustainability*. 2010. Vol. 2, no. 12, p. 3681–3694.

115. DODGE, Samuel and KARAM, Lina. Understanding how image quality affects deep neural networks. In: *2016 Eighth international conference on the quality of multimedia experience (QoMEX)*. IEEE, 2016. p. 1–6. ISBN 1509003541.

116. KOZIARSKI, Michał and CYGANEK, Bogusław. Impact of low resolution on image recognition with deep neural networks: An experimental study. *International Journal of Applied Mathematics and Computer Science*. 2018. Vol. 28, no. 4, p. 735–744.

117. CAI, Dingding, CHEN, Ke, QIAN, Yanlin and KÄMÄRÄINEN, Joni-Kristian. Convolutional low-resolution fine-grained classification. *Pattern Recognition Letters*. 2019. Vol. 119, p. 166–171.

118. EL-ASHMONY, E, EL-DOSUKY, M and ELMOUGY, Samir. Classification of low-quality images using convolutional neural network and deep belief network. *International Journal of Intelligent Computing and Information Sciences*. 2016. Vol. 16, no. 4, p. 19–28.

119. CAP, Quan Huu, UGA, Hiroyuki, KAGIWADA, Satoshi and IYATOMI, Hitoshi. Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering*. 2020. Vol. 19, no. 2, p. 1258–1267.

120. TIAN, Yunong, YANG, Guodong, WANG, Zhe, LI, En and LIANG, Zize. Detection of apple lesions in orchards based on deep learning methods of cyclegan and yolov3-dense. *Journal of Sensors*. 2019. Vol. 2019.

121. DOUARRE, Clément, CRISPIM-JUNIOR, Carlos F, GELIBERT, Anthony, TOUGNE, Laure and ROUSSEAU, David. Novel data augmentation strategies to boost supervised segmentation of plant disease. *Computers and electronics in agriculture*. 2019. Vol. 165, p. 104967.

122. XU, Mingle, YOON, Sook, FUENTES, Alvaro, YANG, Jucheng and PARK, Dong Sun. Style-consistent image translation: A novel data augmentation paradigm to improve plant disease recognition. *Frontiers in Plant Science*. 2022. Vol. 12, p. 3361.

123. KHAN, M Attique, AKRAM, Tallha, SHARIF, Muhammad, JAVED, Kashif, RASHID, Muhammad and BUKHARI, Syed Ahmad Chan. An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Computing and Applications*. 2020. Vol. 32, p. 15929–15948.

124. BRAHIMI, Mohammed, BOUKHALFA, Kamel and MOUSSAOUI, Abdelouahab. Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence*. 2017. Vol. 31, no. 4, p. 299–315.

125. MALŪKAS, Ugnius, MASKELIŪNAS, Rytis, DAMAŠEVIČIUS, Robertas and WOŹNIAK, Marcin. Real-time pathfinding for assisted living using deep learning. *J. Univers. Comput. Sci*. 2018. Vol. 24, no. 4, p. 475–487.

126. YUE-HEI NG, Joe, YANG, Fan and DAVIS, Larry S. Exploiting local features from deep networks for image retrieval. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2015. p. 53–61.

127. GABRYEL, Marcin and DAMAŠEVIČIUS, Robertas. The image classification with different types of image features. In: *Artificial Intelligence and Soft Computing: 16th International Conference, ICAISC 2017, Zakopane, Poland, June 11-15, 2017, Proceedings, Part I 16*. Springer, 2017. p. 497–506. ISBN 3319590626.

128. XING, Shuli, LEE, Marely and LEE, Keun-kwang. Citrus pests and diseases recognition model using weakly dense connected convolution network. *Sensors*. 2019. Vol. 19, no. 14, p. 3195.

129. PICON, Artzai, ALVAREZ-GILA, Aitor, SEITZ, Maximiliam, ORTIZ-BARREDO, Amaia, ECHAZARRA, Jone and JOHANNES, Alexander. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture*. 2019. Vol. 161, p. 280–290.

130. LIANG, Qiaokang, XIANG, Shao, HU, Yucheng, COPPOLA, Gianmarc, ZHANG, Dan and SUN, Wei. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Computers and electronics in agriculture*. 2019. Vol. 157, p. 518–529.

131. ESGARIO, José G M, KROHLING, Renato A and VENTURA, José A. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*. 2020. Vol. 169, p. 105162.

132. ASAD, Muhammad Hamza and BAIS, Abdul. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Information Processing in Agriculture*. 2020. Vol. 7, no. 4, p. 535–545.

133. DARWISH, Ashraf, EZZAT, Dalia and HASSANIEN, Aboul Ella. An optimized

model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant disease diagnosis. *Swarm and evolutionary computation*. 2020. Vol. 52, p. 100616.

134. KAMAL, K C, YIN, Zhendong, WU, Mingyang and WU, Zhilu. Depthwise separable convolution architectures for plant disease classification. *Computers and electronics in agriculture*. 2019. Vol. 165, p. 104948.

135. TOO, Edna Chebet, YUJIAN, Li, NJUKI, Sam and YINGCHUN, Liu. A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*. 2019. Vol. 161, p. 272–279.

136. WOŹNIAK, Marcin and POŁAP, Dawid. An adaptive neuro-heuristic hybrid model for fruit peel defects detection. *Neural Networks*. 2018. Vol. 98, p. 16–33.

137. CHOUHAN, Siddharth Singh, KAUL, Ajay, SINGH, Uday Pratap and JAIN, Sanjeev. Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. *Ieee Access*. 2018. Vol. 6, p. 8852–8863.

138. CAPIZZI, Giacomo, LO SCIUTO, GRAZIA, NAPOLI, Christian, TRAMONTANA, Emiliano and WOŹNIAK, MARCIN. A Novel Neural Networks-Based Texture Image Processing Algorithm for Orange Defects Classification. *International Journal of Computer Science & Applications*. 2016. Vol. 13, no. 2.

139. CODELLA, Noel C F, NGUYEN, Q-B, PANKANTI, Sharath, GUTMAN, David A, HELBA, Brian, HALPERN, Allan C and SMITH, John R. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*. 2017. Vol. 61, no. 4/5, p. 1–5.

140. AHMED, H M, AL-AZAWI, R J and ABDULHAMEED, A A. Evaluation methodology between globalization and localization features approaches for skin cancer lesions classification. In: *Journal of Physics: Conference Series*. IOP Publishing, 2018. p. 012029. ISBN 1742-6596.

141. BARATA, Catarina, RUELA, Margarida, FRANCISCO, Mariana, MENDONÇA, Teresa and MARQUES, Jorge S. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*. 2013. Vol. 8, no. 3, p. 965–979.

142. BRINKER, Titus J, HEKLER, Achim, ENK, Alexander H, BERKING, Carola, HAFERKAMP, Sebastian, HAUSCHILD, Axel, WEICHENTHAL, Michael, KLODE, Joachim, SCHADENDORF, Dirk and HOLLAND-LETZ, Tim. Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer*. 2019. Vol. 119, p. 11–17.

143. RATUL, Md Aminur Rab, MOZAFFARI, M Hamed, LEE, Won-Sook and PARIMBELLI, Enea. Skin lesion classification using deep learning based on dilated convolution. *BioRxiv*. 2019. P. 860700.

144. OH, Byung Ho, KIM, Ki Hean and CHUNG, Kee Yang. Skin imaging using ultrasound imaging, optical coherence tomography, confocal microscopy, and two-photon microscopy in cutaneous oncology. *Frontiers in medicine*. 2019. Vol. 6, p. 274.

145. BARATA, Catarina, CELEBI, M Emre and MARQUES, Jorge S. A survey of feature extraction in dermoscopy image analysis of skin cancer. *IEEE journal of biomedical and health informatics*. 2018. Vol. 23, no. 3, p. 1096–1109.

146. MAITI, Ananjan and CHATTERJEE, Biswajoy. Improving detection of Melanoma and Naevus with deep neural networks. *Multimedia Tools and Applications*. 2020. Vol. 79, no. 21–22, p. 15635–15654.

147. HOSNY, Khalid M, KASSEM, Mohamed A and FOAUD, Mohamed M. Classification of skin lesions using transfer learning and augmentation with Alex-net. *PloS one*. 2019. Vol. 14, no. 5, p. e0217293.

148. PACHECO, Andre G C and KROHLING, Renato A. The impact of patient clinical information on automated skin cancer detection. *Computers in biology and medicine*. 2020. Vol. 116, p. 103545.

149. CURIEL-LEWANDROWSKI, Clara, NOVOA, Roberto A, BERRY, Elizabeth, CELEBI, M Emre, CODELLA, Noel, GIUSTE, Felipe, GUTMAN, David, HALPERN, Allan, LEACHMAN, Sancy and LIU, Yuan. Artificial intelligence approach in melanoma. *Melanoma*. 2019. P. 1–31.

150. KASSANI, Sara Hosseinzadeh and KASSANI, Peyman Hosseinzadeh. A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell*. 2019. Vol. 58, p. 76–83.

151. HEKLER, Achim, UTIKAL, Jochen Sven, ENK, Alexander H, BERKING, Carola, KLODE, Joachim, SCHADENDORF, Dirk, JANSEN, Philipp, FRANKLIN, Cindy, HOLLAND-LETZ, Tim and KRAHL, Dieter. Pathologist-level classification of histopathological melanoma images with deep neural networks. *European Journal of Cancer*. 2019. Vol. 115, p. 79–83.

152. NASR-ESFAHANI, Ebrahim, RAFIEI, Shima, JAFARI, Mohammad H, KARIMI, Nader, WROBEL, James S, SAMAVI, Shadrokh and SOROUSHMEHR, S M Reza. Dense pooling layers in a fully convolutional network for skin lesion segmentation. *Computerized Medical Imaging and Graphics*. 2019. Vol. 78, p. 101658.

153. CHAWLA, Nitesh V, BOWYER, Kevin W, HALL, Lawrence O and KEGELMEYER, W Philip. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002. Vol. 16, p. 321–357.

154. URBONAS, Augustas, RAUDONIS, Vidas, MASKELIŪNAS, Rytis and DAMAŠEVIČIUS, Robertas. Automated identification of wood veneer surface defects using a faster region-based convolutional neural network with data augmentation and transfer learning. *Applied Sciences*. 2019. Vol. 9, no. 22, p. 4898.

155. LEGUEN-DEVARONA, Ireimis, MADERA, Julio, MARTÍNEZ-LÓPEZ, Yoan and HERNÁNDEZ-NIETO, José Carlos. SMOTE-Cov: A New Oversampling Method Based on the Covariance Matrix. In: *Data Analysis and Optimization for Engineering and Computing Problems: Proceedings of the 3rd EAI International Conference on Computer Science and Engineering and Health Services*. Springer, 2020. p. 207–215.

156. HOSNY, Khalid M, KASSEM, Mohamed A and FOAUD, Mohamed M. Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. *Multimedia Tools and Applications*. 2020. Vol. 79, p. 24029–24055.

157. AYAN, Enes and ÜNVER, Halil Murat. Data augmentation is important for the classification of skin lesions via deep learning. In: *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*. IEEE, 2018. p. 1–4. ISBN 1538651351.

158. SHEN, Shuwei, XU, Mengjuan, ZHANG, Fan, SHAO, Pengfei, LIU, Honghong, XU, Liang, ZHANG, Chi, LIU, Peng, YAO, Peng and XU, Ronald X. A low-cost high-performance data augmentation for deep learning-based skin lesion classification. *BME Frontiers*. 2022.

159. BOZKURT, Ferhat. Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach. *Multimedia Tools and Applications*. 2023. Vol. 82, no. 12, p. 18985–19003.

160. PARRA-DOMINGUEZ, Gemma S, GARCIA-CAPULIN, Carlos H and SANCHEZ-YANEZ, Raul E. Automatic facial palsy diagnosis as a classification problem using

regional information extracted from a photograph. *Diagnostics*. 2022. Vol. 12, no. 7, p. 1528.

161. ZHANG, Yuwei, ZHANG, Jun, XIAO, Yi, ZHAO, Ning, WU, Jingyu, WU, Hao, LI, Weifeng, LI, Xin, MA, Ming and SONG, Aiguo. Early Recognition of Facial Paralysis for Rehabilitation of Stroke Patients Using Visual Perception and AI-Assisted Analysis. In: *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2022. p. 25–30. ISBN 1665483067.

162. MCLAUGHLIN, Niall, DEL RINCON, Jesus Martinez and MILLER, Paul. Data-augmentation for reducing dataset bias in person re-identification. In: *2015 12th IEEE International Conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2015. p. 1–6. ISBN 1467376329.

163. LIU, Xin, XIA, Yifan, YU, Hui, DONG, Junyu, JIAN, Muwei and PHAM, Tuan D. Region-based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2020. Vol. 28, no. 10, p. 2325–2332.

164. WANG, Ting, ZHANG, Shu, LIU, Li'An, WU, Gengkun and DONG, Junyu. Automatic facial paralysis evaluation augmented by a cascaded encoder network structure. *IEEE Access*. 2019. Vol. 7, p. 135621–135631.

165. THAI, Huy-Tan, LE, Kim-Hung and NGUYEN, Ngan Luu-Thuy. FormerLeaf: An efficient vision transformer for Cassava Leaf Disease detection. *Computers and Electronics in Agriculture*. 2023. Vol. 204, p. 107518.

166. YU, Sheng, XIE, Li and HUANG, Qilei. Inception convolutional vision transformers for plant disease identification. *Internet of Things*. 2023. Vol. 21, p. 100650.

167. SHARMA, Vivek, TRIPATHI, Ashish Kumar and MITTAL, Himanshu. DLMC-Net: Deeper lightweight multi-class classification model for plant leaf disease detection. *Ecological Informatics*. 2023. Vol. 75, p. 102025.

168. KAUR, Prabhjot, HARNAL, Shilpi, TIWARI, Rajeev, UPADHYAY, Shuchi, BHATIA, Surbhi, MASHAT, Arwa and ALABDALI, Aliaa M. Recognition of leaf disease using the hybrid convolutional neural network by applying feature reduction. *Sensors*. 2022. Vol. 22, no. 2, p. 575.

169. JIANG, Peng, CHEN, Yuehan, LIU, Bin, HE, Dongjian and LIANG, Chunquan. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access*. 2019. Vol. 7, p. 59069–59080.

170. HU, Gensheng, WU, Haoyu, ZHANG, Yan and WAN, Mingzhu. A low shot learning method for tea leaf disease identification. *Computers and Electronics in Agriculture*. 2019. Vol. 163, p. 104852.

171. KARTHIK, R, HARIHARAN, M, ANAND, Sundar, MATHIKSHARA, Priyanka, JOHNSON, Annie and MENAKA, R. Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing*. 2020. Vol. 86, p. 105933.

172. ALRUWAILI, Madallah, ALANAZI, Saad, ABD EL-GHANY, Sameh and SHEHAB, Abdulaziz. An efficient deep learning model for Olive disease detection. *International Journal of Advanced Computer Science and Applications*. 2019. Vol. 10, no. 8.

173. ALWAKID, Ghadah, GOUDA, Walaa, HUMAYUN, Mamoona and SAMA, Najm Us. Melanoma detection using deep learning-based classifications. In: *Healthcare*. MDPI, 2022. p. 2481. ISBN 2227-9032.

174. ALAM, Talha Mahboob, SHAUKAT, Kamran, KHAN, Waseem Ahmad, HAMEED, Ibrahim A, ALMUQREN, Latifah Abd, RAZA, Muhammad Ahsan, ASLAM, Memoona and LUO, Suhuai. An efficient deep learning-based skin cancer classifier

for an imbalanced dataset. *Diagnostics*. 2022. Vol. 12, no. 9, p. 2115.

175. ADEGUN, Adekanmi and VIRIRI, Serestina. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of the state-of-the-art. *Artificial Intelligence Review*. 2021. Vol. 54, p. 811–841.

176. SERTE, Sertan and DEMIREL, Hasan. Gabor wavelet-based deep learning for skin lesion classification. *Computers in biology and medicine*. 2019. Vol. 113, p. 103423.

177. BISLA, Devansh, CHOROMANSKA, Anna, BERMAN, Russell S, STEIN, Jennifer A and POLSKY, David. Towards automated melanoma detection with deep learning: Data purification and augmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2019. p. 0.

178. YU, Zhen, JIANG, Xudong, ZHOU, Feng, QIN, Jing, NI, Dong, CHEN, Siping, LEI, Baiying and WANG, Tianfu. Melanoma recognition in dermoscopy images via aggregated deep convolutional features. *IEEE Transactions on Biomedical Engineering*. 2018. Vol. 66, no. 4, p. 1006–1016.

179. HARANGI, Balazs. Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*. 2018. Vol. 86, p. 25–32.

180. BI, Lei, KIM, Jinman, AHN, Euijoon FENG, Dagan. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. *arXiv preprint arXiv:1703.04197*. 2017.

181. ABAYOMI-ALLI, Olusola O, DAMAŠEVIČIUS, Robertas, ABBASI, Aaqif Afzaal and MASKELIŪNAS, Rytis. Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques. *Electronics*. 2022. Vol. 11, no. 16, p. 2520.

182. ZINEMANAS, Pablo, ROCAMORA, Martín, MIRON, Marius, FONT, Frederic and SERRA, Xavier. An interpretable deep learning model for automatic sound classification. *Electronics*. 2021. Vol. 10, no. 7, p. 850.

183. CROCCO, Marco, CRISTANI, Marco, TRUCCO, Andrea and MURINO, Vittorio. Audio surveillance: A systematic review. *ACM Computing Surveys (CSUR)*. 2016. Vol. 48, no. 4, p. 1–46.

184. AZIMI, Mohammadreza and ROEDIG, Utz. Room identification with personal voice assistants. In: *European Symposium on Research in Computer Security*. Springer, 2021. p. 317–327.

185. KAPOČIŪTĖ-DZIKIENĖ, Jurgita. A domain-specific generative chatbot trained from little data. *Applied Sciences*. 2020. Vol. 10, no. 7, p. 2221.

186. SHAH, Sayed Khushal, TARIQ, Zeenat and LEE, Yugyung. Audio IoT analytics for home automation safety. In the *2018 IEEE International Conference on Big Data (big data)*. IEEE, 2018. p. 5181–5186. ISBN 1538650355.

187. GHOLIZADEH, Samira, LEMAN, Z and BAHARUDIN, BTHT. A review of the application of acoustic emission technique in engineering. *Struct. Eng. Mech*. 2015. Vol. 54, no. 6, p. 1075–1095.

188. HENRIQUEZ, Patricia, ALONSO, Jesus B, FERRER, Miguel A and TRAVIESO, Carlos M. Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on systems, man, and cybernetics: Systems*. 2013. Vol. 44, no. 5, p. 642–652.

189. LOZANO, Héctor, HERNÁEZ, Inmaculada, PICÓN, Artzai, CAMARENA, Javier and NAVAS, Eva. Audio classification techniques in home environments for elderly/dependant people. In: *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July 14-16, 2010. Proceedings 12*. Springer, 2010. p. 320–323. ISBN 3642140963.

190. BEAR, Helen L, HEITTOLA, Toni, MESAROS, Annamaria, BENETOS, Emmanouil

and VIRTANEN, Tuomas. City classification from multiple real-world sound scenes. In: *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019. p. 11–15. ISBN 1728111234.

191. COPETTI CALLAI, Sergio and SANGIORGI, Cesare. A review on acoustic and skid resistance solutions for road pavements. *Infrastructures*. 2021. Vol. 6, no. 3, p. 41.

192. BLUMSTEIN, Daniel T, MENNILL, Daniel J, CLEMINS, Patrick, GIROD, Lewis, YAO, Kung, PATRICELLI, Gail, DEPPE, Jill L, KRAKAUER, Alan H, CLARK, Christopher and CORTOPASSI, Kathryn A. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*. 2011. Vol. 48, no. 3, p. 758–767.

193. BOUNTOURAKIS, Vasileios, VRYSIS, Lazaros and PAPANIKOLAOU, George. Machine learning algorithms for environmental sound recognition: Towards soundscape semantics. In: *Proceedings of the Audio Mostly 2015 on Interaction With Sound*. 2015. p. 1–7.

194. NAJAFABADI, Maryam M, VILLANUSTRE, Flavio, KHOSHGOFTAAR, Taghi M, SELIYA, Naeem, WALD, Randall and MUHAREMAGIC, Edin. Deep learning applications and challenges in big data analytics. *Journal of big data*. 2015. Vol. 2, no. 1, p. 1–21.

195. MARIN, Ivana, KUZMANIC SKELIN, Ana and GRUJIC, Tamara. Empirical evaluation of the effect of optimization and regularization techniques on the generalization performance of the deep convolutional neural network. *Applied Sciences*. 2020. Vol. 10, no. 21, p. 7817.

196. BERGSTRA, James, YAMINS, Daniel and COX, David. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: *International conference on machine learning*. PMLR, 2013. p. 115–123.

197. KHALID, Rabiya and JAVAID, Nadeem. A survey on hyperparameters optimization algorithms of forecasting models in smart grid. *Sustainable Cities and Society*. 2020. Vol. 61, p. 102275.

198. KALLIOLA, Jussi, KAPOČIŪTĖ-DZIKIENĖ, Jurgita and DAMAŠEVIČIUS, Robertas. Neural network hyperparameter optimization for prediction of real estate prices in Helsinki. *PeerJ computer science*. 2021. Vol. 7, p. e444.

199. KOSZEWSKI, Damian and KOSTEK, Bozena. Musical instrument tagging using data augmentation and effective noisy data processing. *Journal of the Audio Engineering Society*. 2020. Vol. 68, no. 1/2, p. 57–65.

200. TAGAWA, Y, MASKELIUNAS, R. and DAMAŠEVICIUS, R. *Acoustic Anomaly Detection of Mechanical Failures in Noisy Real-Life Factory Environments. Electronics 2021, 10, 2329*. 2021.

201. QURTHOBI, Ahmad, MASKELIŪNAS, Rytis and DAMAŠEVIČIUS, Robertas. Detection of Mechanical Failures in Industrial Machines Using Overlapping Acoustic Anomalies: A Systematic Literature Review. *Sensors*. 2022. Vol. 22, no. 10, p. 3888.

202. DOMINGOS, Lucas C F, SANTOS, Paulo E, SKELTON, Phillip S M, BRINKWORTH, Russell S A and SAMMUT, Karl. A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance. *Sensors*. 2022. Vol. 22, no. 6, p. 2181.

203. JI, Chunyan, MUDIYANSELAGE, Thosini Bamunu, GAO, Yutong and PAN, Yi. A review of infant cry analysis and classification. *EURASIP Journal on Audio, Speech, and Music Processing*. 2021. Vol. 2021, no. 1, p. 1–17.

204. SINGH, Krishna Kumar, YU, Hao, SARMASI, Aron, PRADEEP, Gautam and LEE,

Yong Jae. Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. *arXiv preprint arXiv:1811.02545*. 2018.

205. PERVAIZ, Ayesha, HUSSAIN, Fawad, ISRAR, Huma, TAHIR, Muhammad Ali, RAJA, Fawad Riasat, BALOCH, Naveed Khan, ISHMANOV, Farruh and ZIKRIA, Yousaf Bin. Incorporating noise robustness in speech command recognition by noise augmentation of training data. *Sensors*. 2020. Vol. 20, no. 8, p. 2326.

206. MUSHTAQ, Zohaib, SU, Shun-Feng and TRAN, Quoc-Viet. Spectral image-based environmental sound classification using CNN with meaningful data augmentation. *Applied Acoustics*. 2021. Vol. 172, p. 107581.

207. LOEY, Mohamed and MIRJALILI, Seyedali. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. *Computers in biology and medicine*. 2021. Vol. 139, p. 105020.

208. PAHAR, Madhurananda, KLOPPER, Marisa, WARREN, Robin and NIESLER, Thomas. COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. *Computers in biology and medicine*. 2022. Vol. 141, p. 105153.

209. ERDOĞAN, Yunus Emre and NARIN, Ali. COVID-19 detection with traditional and deep features on cough acoustic signals. *Computers in Biology and Medicine*. 2021. Vol. 136, p. 104765.

210. SAIT, Unais, KV, Gokul Lal, SHIVAKUMAR, Sanjana, KUMAR, Tarun, BHAUMIK, Rahul, PRAJAPATI, Sunny, BHALLA, Kriti and CHAKRAPANI, Anaghaa. A deep-learning-based multimodal system for COVID-19 diagnosis using breathing sounds and chest X-ray images. *Applied Soft Computing*. 2021. Vol. 109, p. 107522.

211. SOLTANIAN, Mohammad and BORNA, Keivan. COVID-19 recognition from cough sounds using the lightweight separable-quadratic convolutional network. *Biomedical Signal Processing and Control*. 2022. Vol. 72, p. 103333.

212. MOHAMMED, Emad A, KEYHANI, Mohammad, SANATI-NEZHAD, Amir, HEJAZI, S Hossein and FAR, Behrouz H. An ensemble learning approach to digital coronavirus preliminary screening from cough sounds. *Scientific Reports*. 2021. Vol. 11, no. 1, p. 15404.

213. DESPOTOVIC, Vladimir, ISMAEL, Muhannad, CORNIL, Maël, MC CALL, Roderick and FAGHERAZZI, Guy. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Computers in Biology and Medicine*. 2021. Vol. 138, p. 104944.

214. WANG, Jia-Ching, LEE, Yuan-Shan, LIN, Chang-Hong, SIAHAAN, Ernestasia and YANG, Chung-Hsien. Robust environmental sound recognition with fast noise suppression for home automation. *IEEE Transactions on Automation Science and Engineering*. 2015. Vol. 12, no. 4, p. 1235–1242.

215. STEINFATH, Elsa, PALACIOS-MUÑOZ, Adrian, ROTTSCHÄFER, Julian R, YUEZAK, Deniz and CLEMENS, Jan. Fast and accurate annotation of acoustic signals with deep neural networks. *Elife*. 2021. Vol. 10, p. e68837.

216. TSALERA, Eleni, PAPADAKIS, Andreas and SAMARAKOU, Maria. Comparison of pre-trained CNNs for audio classification using transfer learning. *Journal of Sensor and Actuator Networks*. 2021. Vol. 10, no. 4, p. 72.

217. SALAMON, Justin and BELLO, Juan Pablo. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*. 2017. Vol. 24, no. 3, p. 279–283.

218. NAM, Hyeonuk, KIM, Seong-Hu and PARK, Yong-Hwa. Filteraugment: An acoustic environmental data augmentation method. In: *ICASSP 2022-2022 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022. p. 4308–4312. ISBN 1665405406.

219. MADHU, Aswathy and KUMARASWAMY, Suresh. Data augmentation using the generative adversarial network for environmental sound classification. In: *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019. p. 1–5. ISBN 9082797038.

220. INOUE, Tadanobu, VINAYAVEKHIN, Phongtharin, WANG, Shiqiang, WOOD, David, GRECO, Nancy and TACHIBANA, Ryuki. Domestic activities classification based on CNN using shuffling and mixing data augmentation. *Detection and Classification of Acoustic Scenes and Events 2018 Challenge*. 2018.

221. KHAMPARIA, Aditya, GUPTA, Deepak, DE ALBUQUERQUE, Victor Hugo C, SANGAIAH, Arun Kumar and JHAVERI, Rutvij H. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *The Journal of Supercomputing*. 2020. Vol. 76, p. 8590–8608.

222. SABA, Tanzila, KHAN, Muhammad Attique, REHMAN, Amjad and MARIE-SAINTE, Souad Larabi. Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. *Journal of medical systems*. 2019. Vol. 43, no. 9, p. 289.

223. CALTECH. Caltech Face Database. Online. 1999. [Accessed 3 March 2021]. Available from: http://www.vision.caltech.edu/archive.html

224. ANSARI, Sharik Ali, JERRIPOTHULA, Koteswar Rao, NAGPAL, Pragya and MITTAL, Ankush. Eye-focused Detection of Bell's Palsy in Videos. *arXiv preprint arXiv:2201.11479*. 2022.

225. TAN, Xiangyong, YANG, Jie and CAO, Jiuwen. Facial Nerve Paralysis Assessment based on Regularized Correntropy Criterion SSELM vc and Cascade CNN. In: *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021. p. 1043–1047. ISBN 1665458283.

226. GOGU, Sridhar Reddy and SATHE, Shailesh R. autoFPR: An Efficient Automatic Approach for Facial Paralysis Recognition Using Facial Features. *INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS*. 2023. Vol. 32, no. 02, p. 2340005.

227. LILHORE, Umesh Kumar, IMOIZE, Agbotiname Lucky, LEE, Cheng-Chi, SIMAIYA, Sarita, PANI, Subhendu Kumar, GOYAL, Nitin, KUMAR, Arun and LI, Chun-Ta. Enhanced convolutional neural network model for cassava leaf disease identification and classification. *Mathematics*. 2022. Vol. 10, no. 4, p. 580.

228. C. R, Dhivyaa, KANDASAMY, Nithya and RAJENDRAN, Sudhakar. Integration of dilated convolution with residual dense block network and multi-level feature detection network for cassava plant leaf disease identification. *Concurrency and Computation: Practice and Experience*. 2022. Vol. 34, no. 11, p. e6879.

229. ZHONG, Yiwei, HUANG, Baojin and TANG, Chaowei. Classification of cassava leaf disease based on a non-balanced dataset using transformer-embedded ResNet. *Agriculture*. 2022. Vol. 12, no. 9, p. 1360.

230. ACAR, Emrullah, ERTUGRUL, Omer Faruk, ALDEMIR, Erdogan and OZTEKIN, Abdulkerim. Automatic identification of cassava leaf diseases utilizing morphological hidden patterns and multi-feature textures with a distributed structure-based classification approach. *Journal of Plant Diseases and Protection*. 2022. Vol. 129, no. 3, p. 605–621.

231. ZHANG, Jiayu, QI, Chao, MECHA, Peter, ZUO, Yi, BEN, Zongyou, LIU, Haolu and CHEN, Kunjie. Pseudo-high-frequency boosts the generalization of a convolutional

neural network for cassava disease detection. *Plant Methods*. 2022. Vol. 18, no. 1, p. 136.

232. OYEWOLA, David Opeoluwa, DADA, Emmanuel Gbenga, MISRA, Sanjay and DAMAŠEVIČIUS, Robertas. Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. *PeerJ Computer Science*. 2021. Vol. 7, p. e352.

233. GAJERA, Himanshu K, NAYAK, Deepak Ranjan and ZAVERI, Mukesh A. A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. *Biomedical Signal Processing and Control*. 2023. Vol. 79, p. 104186.

234. BALAHA, Hossam Magdy and HASSAN, Asmaa El-Sayed. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Computing and Applications*. 2023. Vol. 35, no. 1, p. 815–853.

235. ALENEZI, Fayadh, ARMGHAN, Ammar and POLAT, Kemal. Wavelet transform-based deep residual neural network and ReLU-based Extreme Learning Machine for skin lesion classification. *Expert Systems with Applications*. 2023. Vol. 213, p. 119064.

236. BENYAHIA, Samia, MEFTAH, Boudjelal and LÉZORAY, Olivier. Multi-feature extraction based on deep learning for skin lesion classification. *Tissue and Cell*. 2022. Vol. 74, p. 101701.

237. ZAFAR, Kashan, GILANI, Syed Omer, WARIS, Asim, AHMED, Ali, JAMIL, Mohsin, KHAN, Muhammad Nasir and SOHAIL KASHIF, Amer. Skin lesion segmentation from dermoscopic images using convolutional neural network. *Sensors*. 2020. Vol. 20, no. 6, p. 1601.

238. AL-MASNI, Mohammed A, AL-ANTARI, Mugahed A, CHOI, Mun-Taek, HAN, Seung-Moo and KIM, Tae-Seong. Skin lesion segmentation in dermoscopy images via deep full-resolution convolutional networks. *Computer methods and programs in biomedicine*. 2018. Vol. 162, p. 221–231.

239. HASAN, Md Kamrul, DAHAL, Lavsen, SAMARAKOON, Prasad N, TUSHAR, Fakrul Islam and MARTÍ, Robert. DSNet: Automatic dermoscopic skin lesion segmentation. *Computers in biology and medicine*. 2020. Vol. 120, p. 103738.

240. SHARMA, Neeraj, KRISHNAN, Prashant, KUMAR, Rohit, RAMOJI, Shreyas, CHETUPALLI, Srikanth Raj, GHOSH, Prasanta Kumar and GANAPATHY, Sriram. Coswara--a database of breathing, cough, and voice sounds for COVID-19 diagnosis. *arXiv preprint arXiv:2005.10548*. 2020.

241. KUMAR, Santosh, GUPTA, Sachin Kumar, KUMAR, Vinit, KUMAR, Manoj, CHAUBE, Mithilesh Kumar and NAIK, Nenavath Srinivas. Ensemble multimodal deep learning for early diagnosis and accurate classification of COVID-19. *Computers and Electrical Engineering*. 2022. Vol. 103, p. 108396.

242. PAHAR, Madhurananda, KLOPPER, Marisa, WARREN, Robin and NIESLER, Thomas. COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*. 2021. Vol. 135, p. 104572.

243. CHOWDHURY, Nihad Karim, KABIR, Muhammad Ashad, RAHMAN, Md Muhtadir and ISLAM, Sheikh Mohammed Shariful. Machine learning for detecting COVID-19 from cough sounds An ensemble-based MCDM method. *Computers in Biology and Medicine*. 2022. Vol. 145, p. 105405.

244. BROWN, Chloë, CHAUHAN, Jagmohan, GRAMMENOS, Andreas, HAN, Jing, HASTHANASOMBAT, Apinan, SPATHIS, Dimitris, XIA, Tong, CICUTA, Pietro and MASCOLO, Cecilia. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. In: *Proceedings of the 26th ACM SIGKDD*

international conference on knowledge discovery & data mining. 2020. p. 3474–3484.

245. MELEK, Mesut. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. *Neural Computing and Applications*. 2021. Vol. 33, no. 24, p. 17621–17632.

246. LIU, Chengwei, SUI, Xiubao, KUANG, Xiaodong, LIU, Yuan, GU, Guohua and CHEN, Qian. Adaptive contrast enhancement for infrared images based on the neighbourhood conditional histogram. *Remote sensing*. 2019. Vol. 11, no. 11, p. 1381.

247. VIOLA, Paul and JONES, Michael. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Ieee, 2001. p. I–I. ISBN 0769512720.

248. FREUND, Yoav and SCHAPIRE, Robert E. A decision-theoretic generalization of online learning and an application to boosting. *Journal of computer and system sciences*. 1997. Vol. 55, no. 1, p. 119–139.

249. DU, Qiang, FABER, Vance and GUNZBURGER, Max. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM review*. 1999. Vol. 41, no. 4, p. 637–676.

250. IANDOLA, Forrest N, HAN, Song, MOSKEWICZ, Matthew W, ASHRAF, Khalid, DALLY, William J and KEUTZER, Kurt. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv preprint arXiv:1602.07360*. Online. 2016. Available from: http://arxiv.org/abs/1602.07360

251. HENDRYCKS, Dan and DIETTERICH, Thomas. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*. 2019.

252. SANDLER, Mark, HOWARD, Andrew, ZHU, Menglong, ZHMOGINOV, Andrey and CHEN, Liang-Chieh. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. p. 4510–4520.

253. LONG, Yin and LIU, Changhua. Research on deep learning method of crop disease identification. In: *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing*. 2019. p. 1–6.

254. FRANGI, Alejandro F, NIESSEN, Wiro J, VINCKEN, Koen L and VIERGEVER, Max A. Multiscale vessel enhancement filtering. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*. Springer, 1998. p. 130–137. ISBN 3540651365.

255. PATTERSON, Roy D, NIMMO-SMITH, Ian, HOLDSWORTH, John and RICE, Peter. An efficient auditory filterbank based on the gammatone function. In: *a meeting of the IOC Speech Group on Auditory Modelling at RSRE*. 1987.

256. BRAUN, Stefan, NEIL, Daniel and LIU, Shih-Chii. A curriculum learning method for improved noise robustness in automatic speech recognition. In: *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017. p. 548–552. ISBN 0992862671.

257. BONCELET, Charles. Image noise models. In: *The essential guide to image processing*. Elsevier, 2009. p. 143–167.

258. VAN DER MAATEN, Laurens and HINTON, Geoffrey. Visualizing data using t-SNE. journal of Machine Learning Research 9. *Nov (2008)*. 2008.

259. KINGMA, Diederik P and BA, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 2014.

166

260. NUSRAT, Ismoilov and JANG, Sung-Bong. A comparison of regularization techniques in deep neural networks. *Symmetry*. 2018. Vol. 10, no. 11, p. 648.

261. WARDE-FARLEY, David, GOODFELLOW, Ian J, COURVILLE, Aaron and BENGIO, Yoshua. An empirical analysis of dropout in piecewise linear networks. *arXiv preprint arXiv:1312.6197*. 2013.

262. HU, Kai, NIU, Xiaorui, LIU, Si, ZHANG, Yuan, CAO, Chunhong, XIAO, Fen, YANG, Wanchun and GAO, Xieping. Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. *Biomedical Signal Processing and Control*. 2019. Vol. 51, p. 200–209.

263. ADJED, Faouzi, SAFDAR GARDEZI, Syed Jamal, ABABSA, Fakhreddine, FAYE, Ibrahima and CHANDRA DASS, Sarat. Fusion of structural and textural features for melanoma recognition. *IET Computer Vision*. 2018. Vol. 12, no. 2, p. 185–195.

264. ELTAYEF, Khalid, LI, Yongmin and LIU, Xiaohui. Detection of melanoma skin cancer in dermoscopy images. In: *Journal of Physics: Conference Series*. IOP Publishing, 2017. p. 012034. ISBN 1742-6596.

265. SULTANA, Nazneen N, MANDAL, Bappaditya and PUHAN, Niladri B. Deep residual network with regularised fisher framework for detection of melanoma. *IET Computer Vision*. 2018. Vol. 12, no. 8, p. 1096–1104.

266. ESMAEILPOUR, Mohammad, CARDINAL, Patrick and KOERICH, Alessandro Lameiras. Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network. *Applied Soft Computing*. 2020. Vol. 86, p. 105912.

267. ALY, Mahmoud, RAHOUMA, Kamel H and RAMZY, Safwat M. Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings. *Alexandria Engineering Journal*. 2022. Vol. 61, no. 5, p. 3487–3500.

268. CHAUDHARI, Gunvant, JIANG, Xinyi, FAKHRY, Ahmed, HAN, Asriel, XIAO, Jaclyn, SHEN, Sabrina and KHANZADA, Amil. Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. *arXiv preprint arXiv:2011.13320*. 2020.

269. BAGAD, Piyush, DALMIA, Aman, DOSHI, Jigar, NAGRANI, Arsha, BHAMARE, Parag, MAHALE, Amrita, RANE, Saurabh, AGARWAL, Neeraj and PANICKER, Rahul. Cough against COVID-19: Evidence of COVID-19 signature in cough sounds. *arXiv preprint arXiv:2009.08790*. 2020.

270. RUNKLER, Thomas A. *Data analytics*. Springer, 2020.

271. ABAYOMI-ALLI, Olusola Oluwakemi, DAMASEVICIUS, Robertas, MISRA, Sanjay, MASKELIUNAS, Rytis and ABAYOMI-ALLI, Adebayo. Malignant skin melanoma detection using image augmentation by oversampling nonlinear lower-dimensional embedding manifold. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2021. Vol. 29, no. 8, p. 2600–2614.

PUBLICATION OF RESEARCH RESULTS

List of Scientific Publications on Dissertation Topics

Articles Indexed in the Web of Science with Impact Factor

1. **Abayomi-Alli, Olusola O.**, Damaševičius, Robertas, Misra, Sanjay, & Abayomi-Alli, Adebayo. (2023). FruitQ: a new dataset of multiple fruit images for freshness evaluation. //*Multimedia Tools and Applications*: Springer. ISSN: 1380-7501, 2023. DOI: 10.1007/s11042-023-16058-6. Science Citation Index Expanded (Web of Science); Scopus; DOAJ) (IF: 3.866; Q1 (2022, InCites JCR SCIE)) (CiteScore: 6.10; SNIP: 1.182; SJR: 0.720; Q1 (2022, Scopus Sources)) (FOR: T 007) (Input: 0.200)

2. **Abayomi-Alli, Olusola O.**, Damaševičius, Robertas, Abbasi, Aaqif Afzaal, & Maskeliūnas, Rytis. (2022). Detection of COVID-19 from deep breathing sounds using sound spectrum with image augmentation and deep learning techniques. // *Electronics*. Basel: MDPI. ISSN 2079-9292. 2022, vol. 11, Iss. 16, art. No. 2520, p. 1–19. DOI: 10.3390/electronics11162520. (Science Citation Index Expanded (Web of Science); Scopus; DOAJ) (IF: 2.690; IF/AIF: 0.61; Q3 (2021, InCites JCR SCIE)) (CiteScore: 3.70; SNIP: 1.013; SJR: 0.590; Q2 (2021, Scopus Sources)) (FOR: T 007) (Input: 0.250)

3. **Abayomi-Alli, Olusola O.**, Damaševičius, Robertas, Maskeliūnas, Rytis, & Misra, Sanjay. (2022). An ensemble learning model for COVID-19 detection from blood test samples. // *Sensors*: MDPI. ISSN 1424-8220. 2022, vol. 22, iss. 6, art. No. 2224, p. 1–18. DOI: 10.3390/s22062224. (Science Citation Index Expanded (Web of Science); Scopus; DOAJ) (IF: 3.847; AIF: 4.050; IF/AIF: 0.90; Q2 (2021, InCites JCR SCIE)) (CiteScore: 6.40; SNIP: 1.420; SJR: 0.803; Q2 (2021, Scopus Sources)) (FOR: T 007) (Input: 0.250)

4. **Abayomi-Alli, O. O.**, Damaševičius, R., Qazi, A., Adedoyin-Olowe, M., & Misra, S. (2022). Data Augmentation and Deep Learning Methods in Sound Classification: A Systematic Review. *Electronics*, 11(22), 3795. *Electronics*. Basel: MDPI. 2022, vol. 11, Iss. 22, art. No. 3795, p. 1–32. DOI: 10.3390/electronics11223795. (Science Citation Index Expanded (Web of Science); Scopus; DOAJ) (IF: 2.690; IF/AIF: 0.61; Q3 (2021, InCites JCR SCIE)) (CiteScore: 3.70; SNIP: 1.013; SJR: 0.590; Q2 (2021, Scopus Sources)) (FOR: T 007) (Input: 0.250)

5. **Abayomi-Alli, Olusola Oluwakemi;** Damaševičius, Robertas; Misra, Sanjay; Maskeliūnas, Rytis. Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning // *Expert Systems*. Hoboken, NJ: John Wiley & Sons. ISSN 0266-4720. eISSN 1468-0394. 2021, vol. 38, iss. 7, e12746, p. 1–21. DOI: 10.1111/exsy.12746. (Science Citation Index Expanded (Web of Science); Scopus) (IF: 2.587; AIF: 4.175; IF/AIF: 0.619; Q2 (2020, InCites JCR

SCIE)) (CiteScore: 3.40; SNIP: 0.900; SJR: 0.365; Q2 (2020, Scopus Sources)) (FOR: T 007) (Input: 0.250)

6. **Abayomi-Alli, Olusola Oluwakemi**; Damaševičius, Robertas; Maskeliūnas, Rytis; Misra, Sanjay. Few-shot learning with a novel Voronoi tessellation-based image augmentation method for facial palsy detection // *Electronics*. Basel: MDPI. ISSN 2079-9292. 2021, vol. 10, iss. 8, art. No. 978, p. 1–18. DOI: 10.3390/electronics10080978. (Science Citation Index Expanded (Web of Science); Scopus; DOAJ) (IF: 2.397; AIF: 4.117; IF/AIF: 0.582; Q3 (2020, InCites JCR SCIE)) (CiteScore: 2.70; SNIP: 0.985; SJR: 0.360; Q2 (2020, Scopus Sources)) (FOR: T 007) (Input: 0.250)

7. **Abayomi-Alli, Olusola Oluwakemi**; Damaševičius, Robertas; Misra, Sanjay; Maskeliūnas, Rytis; Abayomi-Alli, Adebayo. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold // *Turkish Journal of Electrical Engineering and Computer Sciences*. Ankara: TÜBİTAK. ISSN 1300-0632. eISSN 1303-6203. 2021, vol. 29, no. SI-1, p. 2600–2614. DOI: 10.3906/elk-2101-133. (Science Citation Index Expanded (Web of Science); Scopus) (IF: 0.806; AIF: 4.332; IF/AIF: 0.186; Q4 (2020, InCites JCR SCIE)) (CiteScore: 1.80; SNIP: 0.572; SJR: 0.225; Q3 (2020, Scopus Sources)) (FOR: T 007) (Input: 0.200).

List of Other International Scientific Publications

Articles in conference proceedings

1. **Abayomi-Alli, O. O.,** Damaševičius, R., Maskeliūnas, R., & Abayomi-Alli, A. (2020, September). BiLSTM with data augmentation using interpolation methods to improve early detection of Parkinson's disease. In *Proceedings of the 2020 federated conference on computer science and information systems (FedCSIS)*, 6–9 September 2020, Sofia, Bulgaria / M. Ganzha, L. Maciaszek, M. Paprzycki (eds.). Piscataway, NJ: IEEE, 2020. p. 371–380.

2. **Abayomi-Alli, O. O.,** Damaševičius, R., Wieczorek, M., & Woźniak, M. (2020). Data augmentation using principal component resampling for image recognition by deep learning. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020*, Zakopane, Poland, October 12–14, 2020, Proceedings, Part II 19 (p. 39–48). Springer International Publishing.

3. **Abayomi-Alli, O. O.,** Sidekerskienė, T., Damaševičius, R., Siłka, J., & Połap, D. (2020). Empirical Mode Decomposition Based Data Augmentation for Time Series Prediction Using NARX Network. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020*, Zakopane, Poland, October 12–14, 2020, Proceedings, Part I 19 (p. 702–711). Springer International Publishing

OLUSOLA O. ABAYOMI-ALLI
Ph.D.(in-view), M.Sc., B.Sc.

Proficient in conducting experiments with proven experience in research and analytical result interpretation. Detailed-oriented team player, hardworking, excellent organization skills, self-motivated and passionate about research excellence. Skilled in conceptualization, design and implementation of Artificial Intelligence methods for pattern recognition and prediction projects.

CONTACT
Email:
oluaba@ktu.edu
abayomialliolusola@gmail.com

SKILLS
Teaching, Research
Scholar Writings, Supervision
Public Speaking, Leadership
Mentorship, Effective Communication

PROGRAMMING SKILLS
MATLAB, Python, C++,
MySQL TensorFlow, PowerBI

HOBBIES
Travelling
Public speaking
Reading and Dancing

EDUCATION

Ph.D. (Informatics Engineering)
Kaunas University of Technology, Lithuania
November 2019 – To be completed June 2024
Dissertation Title: Small Data Analytics Using Artificial Intelligence Methods

Master's of Science (Computer Science)
Federal University of Agriculture, Abeokuta, Nigeria
Dec 2012 – October, 2015
Dissertation Title: An Extended Short Message Spam Filtering Scheme Using Artificial Immune System

WORK EXPERIENCE

Kaunas University of Technology, Lithuania (Ph.D. Candidate)
Department of Software Engineering
November 2019 – To be completed June 2024
Responsibilities:
- Conducting independent or supporting research on a range of datasets like images, sound, and numerical biomedical, following correct protocols for study integrity.
- Compiling and synthesizing data to identify trends and patterns in disease classification impacting computer vision, sound, and biomedical tasks.
- Adapting advanced statistical approaches and software tools to analyze data to conclude overall outcomes.
- Preparing research writing and experimental results for publication in journals
- Collaboration with other researchers in Artificial Intelligence methods.
- Writing and publishing in high-impact research.

Covenant University Ota – Nigeria (Lecturer II)
Department of Electrical and Computer Engineering
October 2016 – January 2020
Responsibilities:
- Teaching and examining undergraduate Engineering courses with an average of over 600 students from the Engineering Faculty.
- Applied Computer Programming I (GEC 215)
- Applied Computer Programming II (GEC225)
- Object Oriented Design and Programming (CEN 414)
- Technical Communication (GEC 324)
- Computer Network and Security (CEN 523)

- Academic mentor and adviser to students to explore career pathways for student Internship-based and enhanced knowledge, skills and strengths.
- Supported engineering students in practical lab work (Applied programming language) as well as classroom lessons to bridge gaps between theoretical and practical concepts.
- Evaluated and supervised student activities and performance levels to provide reports on academic progress.
- Participated in various campus membership groups to promote programming skills to STEM students.
- Collaborated with faculty and staff to create meaningful learning experiences.
- Participated in conferences and professional development opportunities to stay updated with recent tools in computer science and engineering fields.
- Other responsibilities as assigned by the Head of Departments.

WEBSITES, PORTFOLIO, PROFILES

- **ResearchGate:**
  https://www.researchgate.net/profile/Olusola-Abayomi-Alli
- **LinkedIn:**
  https://www.linkedin.com/in/olusola-abayomi-alli-39b72b130/
- **Google   Scholar:**
  https://scholar.google.com/citations?hl=lt&user=L6en5dcAAAAJ

AWARDS

- **Recipient:** Most Active PhD student, 2022 (amount: 900 euros);
- **Recipient:** Lithuanian Science Council Scholarship (LMT), 2022 (amount: 2,700 euros);
- **Recipient:** Most Active PhD student, 2021 (amount: 900 euros);
- **Recipient:** Lithuanian Science Council Scholarship (LMT), 2021 (amount: 2,980 euros).

REFERENCES

- References are available upon request.

ACKNOWLEDGEMENTS

Finally, I am eternally grateful to my husband and burden bearer, Dr. Adebayo Abayomi-Alli, you are the wheel that kept me going. Thank you for all the editorial roles and the immense contributions in proofreading my dissertation. Thank you, baby, for cheering me up daily, all the late-night jokes and moral support just to keep me entertained and focused as I try to meet up with all the deadlines. Your display of confidence and trust in my abilities was a sufficient driving force for me. Thank you for never complaining despite single-handedly raising our kids in these last four years and doing all the school runs alone. Thank you, my Sunshine and King, God must be deeply in love with me for granting me such a blessing of 'You'.

My heartfelt gratitude to my bundles of joy: 'Precious', 'Audrey', and 'April'. You are a gift of God to me and thank you for showing me nothing but love and understanding all through my studies. I appreciate your sacrifices as I missed all your birthdays, school events and recitals, vacations, transitioning to teenagers, and the list goes on. You are the best kids in the world, I could not have asked God for anything more. Words can't describe how much I love and appreciate you all "My priceless jewels." Thank you for brightening my life. Hugs and kisses to you all.

Thank you, Jesus, for your promises are Ye and Amen. "They that wait upon the Lord shall be strong and shall do great exploits." You are my forever shield, buckler, comforter and my God.

174