PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
M'HAMED BOUGARA-BOUMERDES UNIVERSITY

Faculty of Technology
Electrical Systems Engineering Department
Academic Master Dissertation

Domain: Science and Technology
Field: BIOMEDICAL ENGINEERING
Specialty: BIOMEDICAL INSTRUMENTATION

Intelligent System for Assist of Deafness and Muteness People

Presented by:

- ALKURD Mahmoud
- HALIMI Weam

Defended on: 12 July 2023 before the jury :

Dr. OMARI Taher	MCA	UMBB	President
Dr. BENDAHMANE Bouteina	Doctor	UMBB	Examiner
Pr. AMMAR Mohammed	MCA	UMBB	supervisor
Mr. GUENANI Moussa	Doctoral student	UMBB	Co-supervisor

2022/2023

بسم الله الرحمن الرحيم

Dedication:

I dedicate this work to those who are dearest to me in the world,
To those without whom I would not be here today,
To those who have encouraged and pushed me to go all the way,
To my father and mother, may God keep them with me for as long as I live,
To my **father**, my first supporter, and the source of my strength, **Abdullah Ismail Al-Kurd**, may God protect and preserve him
To my **mother**, to the spring of tenderness, whose prayers continue to accompany me wherever I go and travel, may God protect and preserve him.
To my grandfather **Ismail Al-Kurd Abu Muhammad** and grandmother **Aziza Al-Kurd Umm Muhammad**, may God have mercy on them,
To my grandmother, **Fatima Al-Kurd Umm Atef**, may God protect and preserve her,
To my aunts and uncles who have always been and continue to support me,
To my **uncle** and second father in Alienation, **Abu Tamer Al-Kurd**, may God protect and preserve him
To my honorable family (KURD) in the homeland and diaspora
To my country, my homeland, and my birthplace, Palestine
To my country and my second home, Algeria
To my dear friend, Dr. Guenani Moussa, who had a great credit for accomplishing this work,
To my dear friends and brothers,
to my best friends,
For their love, encouragement and constant support for me, may God grant you success
And finally, to all the people who supported and encouraged me,
Thank you for everything,

Dedication:

To all those who have been my sources of inspiration and support throughout this journey,

I dedicate this dissertation to all of you. Your trust, encouragement, and belief in my abilities have been the foundation of my success. Your words of encouragement have ignited the flame of my passion and pushed me to surpass my limits.

To myparents :

My Mum *DJAMILA*, my Dad *KAMEL*

my Sis *HADJER*

And to all my familly :

who has been an endless source of support, I am deeply grateful. Your unconditional love and infinite patience have been my driving force.

To my friends :

Mahmoud, Nada, Iman, Maroua, Fatima, Chaima, Sabrina,Aisha And Chaimaet ahmed

thank you for standing by my side at every step. Your presence, encouragement, and inspiring conversations have made this academic journey even more meaningful.

Lastly

to all those who will read this dissertation, I hope that this research can contribute in some way to the advancement of science and the promotion of inclusivity for the deaf and mute community.

Thank you all for your unwavering support.

WEAM

Acknowledgement:

We thank the almighty **ALLAH** for granting us the strength and patience to complete this work

We would like to express our deepest gratitude to

Mr. AMMAR Mouhamed, Mr. GEUNANI Moussa, **Mr. MASSAOUDI Noureddine**, **Mme. AISSAOUI Siham**

For taking the responsibility of supervising us, guiding, and advising us throughout the realization of this work, as well as for the trust they have placed in us.

Thanks for all the responsible of the young deaf people school of Rouiba

The director **Mme. SADAOUI Nassira**, its general supervisor
Mr. BOUKHATAYA Moustafa, **Mme. GENAN Souad**
and all of students of the school .

we sincerely thank the members of the jury for the honor they have bestowed upon us by agreeing to be the judges of our thesis.

We extend our heartfelt thanks to those who have provided us with guidance, support and assistance, directly or indirectly, in the realization and implementation of this work

Abstract:

This dissertation focuses on hand gesture recognition, specifically targeting the sign language alphabet. The primary objective is to establish the groundwork for the development of a sign language recognition technology utilizing artificial intelligence. To achieve this, deep learning algorithms such as convolutional neural networks (CNN) were employed.

By employing these learning methods, the accuracy of hand gesture recognition has significantly improved, thereby facilitating easier communication and interaction with the environment for deaf and mute individuals. This advancement aims to enhance their integration into society by providing them with the means to communicate more effectively and adapt more easily to their surroundings, leveraging the latest advancements in artificial intelligence. The ultimate goal of this research is to develop a finalized product that enables seamless communication for this community, empowering them to navigate their daily lives with greater ease.

Keywords:

Deafness and Muteness, Artificial Intelligence, Computer Vision, Deep Learning, Sign Language, CNN.

Resume:

Cette thèse porte sur la reconnaissance des gestes de la main, ciblant spécifiquement l'alphabet de la langue des signes. L'objectif principal est d'établir les bases du développement d'une technologie de reconnaissance de la langue des signes utilisant l'intelligence artificielle. Pour y parvenir, des algorithmes d'apprentissage en profondeur tels que les réseaux de neurones convolutifs (CNN) ont été utilisés.

En utilisant ces méthodes d'apprentissage, la précision de la reconnaissance des gestes de la main s'est considérablement améliorée, facilitant ainsi la communication et l'interaction avec l'environnement pour les personnes sourdes et muettes. Cette avancée vise à favoriser leur intégration dans la société en leur donnant les moyens de mieux communiquer et de s'adapter plus facilement à leur environnement, en s'appuyant sur les dernières avancées de l'intelligence artificielle. Le but ultime de cette recherche est de développer un produit finalisé qui permette une communication transparente pour cette communauté, lui permettant de naviguer plus facilement dans sa vie quotidienne.

Mots clés:

Surdité et mutisme, intelligence artificielle, vision par ordinateur, apprentissage en profondeur, langue des signes, CNN.

ملخص:

تركز هذه الرسالة على التعرف على إيماءات اليد ، وتستهدف على وجه التحديد أبجدية لغة الإشارة. الهدف الأساسي هو إرساء الأساس لتطوير تقنية التعرف على لغة الإشارة باستخدام الذكاء الاصطناعي. لتحقيق ذلك ، تم استخدام خوارزميات التعلم العميق مثل الشبكات العصبية التلافيفية (CNN).

من خلال استخدام أساليب التعلم هذه ، تحسنت دقة التعرف على إيماءات اليد بشكل كبير ، مما يسهل التواصل والتفاعل مع البيئة للأفراد الصم والبكم. يهدف هذا التقدم إلى تعزيز اندماجهم في المجتمع من خلال تزويدهم بوسائل التواصل بشكل أكثر فعالية والتكيف بسهولة أكبر مع محيطهم ، والاستفادة من أحدث التطورات في الذكاء الاصطناعي. الهدف النهائي من هذا البحث هو تطوير منتج نهائي يتيح التواصل السلس لهذا المجتمع ، وتمكينهم من التنقل في حياتهم اليومية بسهولة أكبر.

كلمات مفتاحية:

الصم والبكم، الذكاء الاصطناعي، الرؤية الحاسوبية، التعلم العميق، لغة الاشارة، بنية CNN.

Table of Contents

Table of Contents:

List of Figures

List of Figures :

List of Tables

List of Tables:

List of Abbreviations

List of Abbreviations:

dB: deciBel

WHO: World Health Organization

USD: United States Dollar

BC: Before Christ

AI: Artificial Intelligence

CV: Computer Vision

ML: Machine Learning

DL: Deep Learning

CNN: Convolutional Neural Network

K-NN: K-Nearest Neighbors

SVM: Support Vector Machine

ANN: Artificial Neural Network

MLP: Multilayer Perceptron

RNN: Recurrent Neural Network

BPTT: BackPropagation Through Time

LSTM: Long Short-Term Memory

GRU: Gated Recurrent Unit

FC: Fully Connected

2D: 2 Dimensional

ReLU: Rectified Linear Unit

GPU: Graphics Processing Unit

CPU: Central Processing Unit

ILSVRC: Image Net Large Scale Visual Recognition Challenge

RAM: Random Access Memory

VRAM: Video RAM

GHz: GigaHertz

LTS: Long Term Support

SGD: Stochastic gradient descent

ADAM: Adaptive moment estimation

RMSprop: Root Mean Square Propagation

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

General Introduction

General Introduction:

In recent years, the development of intelligent systems and advanced technologies has opened up new possibilities for assisting individuals with deafness and muteness. These conditions pose unique challenges to communication and engagement, impacting the quality of life for affected individuals. In response, researchers and engineers have focused on leveraging artificial intelligence (AI) and computer vision to create intelligent systems that can support and enhance the communication abilities of deaf and mute individuals.

The objective of this thesis, titled "Intelligent System for Assist of Deafness and Muteness People," is to explore the intersection of deafness, muteness, artificial intelligence, and computer vision, and develop a comprehensive understanding of how these fields can be integrated to create innovative solutions. The thesis will consist of three main chapters: Deafness and Muteness, Artificial Intelligence and Computer Vision, and Implementation, Results, and Discussion.

Chapter I: Deafness and Muteness

The first chapter provides an introduction to deafness and muteness, focusing on the anatomy of the ear and the hearing process. It explores the three main parts of the ear: the outer ear, middle ear, and inner ear, and discusses how they contribute to the hearing process. The chapter also delves into the different types of deafness, historical backgrounds, symptoms, etiology, and medical causes. Similarly, it examines muteness, its introduction, historical background, types, symptoms, etiology, and medical causes. Furthermore, communication methods and the characteristics of sign language are discussed.

Chapter II: Artificial intelligence and Computer Vision

The second chapter introduces computer vision and its relationship with artificial intelligence (AI). It provides an overview of AI, focusing on machine learning and deep learning. The distinction between supervised and unsupervised classification methods is explained, and specific classification techniques such as the K-Nearest Neighbors (K-NN) classifier, decision tree, and support vector machine (SVM) are presented. The chapter also explores artificial neural networks (ANNs), including their biological basis and architecture, and highlights the different types of ANNs, such as perceptrons and multilayer perceptrons (MLPs). Convolutional Neural Networks (CNNs) are extensively discussed, including their architecture, training process, and various CNN models like VGGNet, ResNet, Xception, and MobileNet.

Chapter III: Implementation, Results, and Discussion

The third chapter focuses on the implementation of the intelligent system designed to assist individuals with deafness and muteness. It describes the specific configuration used, including the computer hardware and programming language employed. The datasets used for training and evaluation purposes are presented, with examples such as the Sign Language Digits Dataset, Sign Language Arabic Dataset, and Sign Language French Dataset. The chapter also covers data pre-processing techniques and introduces the CNN models utilized. Training stages and the evaluation of the models using the confusion matrix are explained. The chapter concludes with a discussion of the results obtained, highlighting the performance of the models with respect to different datasets and potential areas of improvement.

Chapter I: Deafness and muteness

Chapter I : Deafness and muteness

1. Introduction:

The human ear is an intricate and remarkable organ responsible for our ability to perceive sound and maintain our balance. Understanding its anatomy and functioning is crucial in comprehending various auditory conditions, such as deafness and muteness. Additionally, the existence of sign language as a unique form of communication provides an alternative means of expression for individuals with hearing or speech impairments. In this introduction, we will provide an overview of the anatomy and functioning of the ear, delve into the topics of deafness and muteness, explore the realm of sign language communication, and gain insights into the captivating world of sign language.

To begin, we will explore the anatomy of the ear, which comprises three main parts: the outer ear, the middle ear, and the inner ear. Each component plays a vital role in capturing, transmitting, and processing sound waves, ultimately enabling us to perceive auditory stimuli. By understanding the intricate structures and their functions within the ear, we can gain a deeper appreciation of how sound is processed and transmitted within the auditory system.

Deafness and muteness, though distinct conditions, are interconnected aspects of human communication. Deafness refers to the partial or complete loss of hearing, which can have various causes and manifestations. We will delve into the types of deafness, explore their symptoms, and investigate the underlying etiology and medical causes associated with this condition. Muteness, on the other hand, refers to the inability to speak or difficulty in verbal communication. We will examine the historical background, symptoms, types, and medical causes of muteness, shedding light on the challenges faced by individuals affected by this condition.

Sign language communication provides a fascinating alternative to oral communication for individuals with hearing or speech impairments. It is a visual-gestural language that utilizes hand movements, facial expressions, and body postures to convey meaning and express thoughts and ideas. We will explore the concept of communication itself, the role of gestures, and delve into the advantages and limitations of sign language communication. By understanding the unique aspects of sign language, we can recognize its importance in facilitating effective communication and promoting inclusivity.

Finally, we will embark on a journey into the world of sign language itself. Sign language is not a universal language but rather a diverse collection of languages that have evolved within different linguistic communities. We will delve into its historical background, exploring the development and recognition of various sign languages worldwide. Furthermore, we will explore the field of semiology, which encompasses the study of signs and their linguistic characteristics within sign languages. By examining some key features and characteristics of sign language, we can appreciate its rich linguistic nature and unique cultural significance.

2. Anatomy of the Ear & Hearing Process:

2.1. Anatomy of the Ear :

The ear is a remarkable organ that allows us to perceive and interpret the world through sound. From the gentle rustling of leaves to the melodious symphonies of a full orchestra, the intricate anatomy of the ear enables us to experience the beauty of auditory sensations. Let us take a closer look at the anatomy of this extraordinary sensory system and understand how it functions [1].

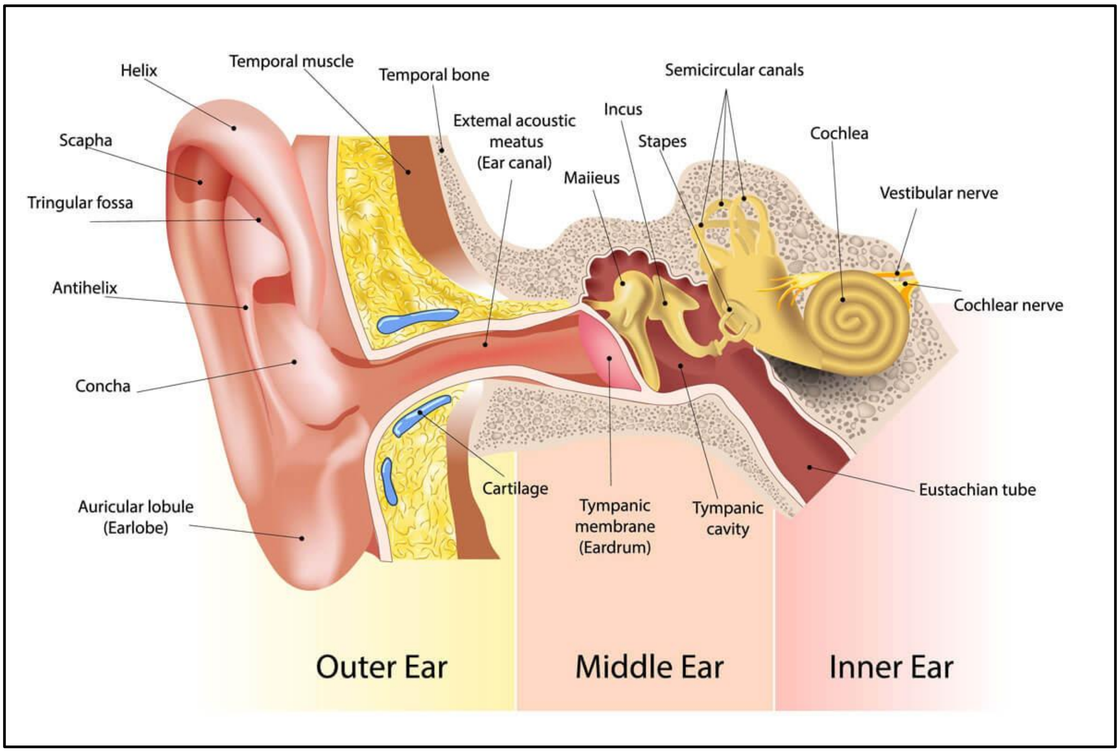

Fig 1. Anatomy of the Ear.

2.2. The ear can be divided into three main parts:

2.2.1. Outer Ear:

The outer ear is the visible part of the ear that includes the pinna, also known as the auricle, and the ear canal. The pinna is the fleshy, cartilaginous structure that protrudes from the side of our heads. Its unique shape helps in capturing sound waves from the surrounding environment and funnels them into the ear canal. The ear canal is a narrow, curved passage that leads to the middle ear.

2.2.2. Middle Ear:

The middle ear is an air-filled cavity located between the eardrum and the inner ear. It consists of three tiny bones called the ossicles: the malleus (hammer), incus (anvil), and stapes (stirrup). These bones form a chain-like structure that efficiently transmits sound vibrations from the eardrum to the inner ear.

When sound waves enter the ear canal, they cause the eardrum to vibrate. These vibrations are then transferred to the malleus, which is connected to the eardrum. The malleus, in turn, passes the vibrations to the incus and then to the stapes. The stapes, the last bone in the chain, is connected to a membrane called the oval window, which separates the middle ear from the inner ear.

2.2.3. Inner Ear:

The inner ear is a complex structure responsible for converting sound vibrations into electrical signals that can be interpreted by the brain.

It consists of two primary components:

- The cochlea, often referred to as the spiral-shaped snail shell, is the main sensory organ for hearing. It is filled with fluid and contains thousands of tiny hair cells, which are the receptors for sound. When the vibrations from the middle ear reach the cochlea, they cause the fluid inside to move, stimulating the hair cells. These hair cells convert the mechanical energy of the vibrations into electrical signals that can be transmitted to the brain via the auditory nerve.
- The vestibular system, located next to the cochlea, is responsible for our sense of balance and spatial orientation. It consists of three semicircular canals and the otolithic organs, which detect rotational movements and linear accelerations, respectively. The vestibular system works in conjunction with the visual and proprioceptive systems to maintain our equilibrium.

2.3. Hearing Process:

Now that we have explored the anatomy of the ear, It can be summarized as follows:

- Sound waves are captured by the pinna and directed into the ear canal.
- The eardrum vibrates in response to the sound waves.
- The ossicles amplify and transmit these vibrations to the oval window.
- The movement of the oval window creates fluid waves inside the cochlea.
- The hair cells in the cochlea are stimulated by the fluid waves.
- The hair cells convert the mechanical energy into electrical signals.
- The electrical signals are transmitted to the brain via the auditory nerve.
- The brain processes and interprets these signals as sound.

3. Deafness and Muteness:

3.1. Deafness:

3.1.1. Introduction:

A person is said to have hearing loss if they are not able to hear as well as someone with normal hearing, meaning hearing thresholds of 20 dB or better in both ears. It can be mild, moderate, moderately severe, severe or profound, and can affect one or both ears. Major causes of hearing loss include congenital or early onset childhood hearing loss, chronic middle ear infections, noise-induced hearing loss, age-related hearing loss, and ototoxic drugs that damage the inner ear. [2]

The impacts of hearing loss are broad and can be profound. They include a loss of the ability to communicate with others delayed language development in children, which can lead to social isolation, loneliness and frustration, particularly among older people with hearing loss. Many areas lack sufficient accommodations for hearing loss, which effect academic performance and options for employment. Children with hearing loss and deafness in developing countries rarely receive any schooling. The World Health Organization (WHO) estimates that unaddressed hearing loss costs the global economy US$ 980 billion annually due to health sector costs (excluding the cost of hearing devices), costs of educational support, loss of productivity and societal costs.

3.1.2. Types of Deafness:

Deafness can be categorized into three main types [3]:

- **Conductive deafness**: occurs when sound waves are unable to reach the inner ear due to issues in the outer or middle ear.
- **Sensorineural deafness**: results from damage or dysfunction in the inner ear or auditory nerve.
- **Mixed deafness**: is a combination of both conductive and sensorineural hearing loss.

3.1.3. Historical Backgroundof Deafness:

Deafness has a long history, with ancient civilizations associating it with divine punishment. In the Middle Ages, it was linked to superstitions and seen as evil. The Renaissance brought recognition of deaf individuals' intellectual abilities. The 18th and 19th centuries saw the establishment of public schools and the emergence of sign language.

The 20th century witnessed advancements in audiology and technology, including hearing aids and cochlear implants. Sign language gained recognition as a linguistic system, and diverse educational approaches emerged. Deafness has evolved from stigma to acceptance, highlighting the importance of embracing communication diversity and empowering deaf individuals in a predominantly hearing-centric world.

3.1.4.Symptoms and Etiology:

Deafness is characterized by partial or complete hearing loss, with symptoms including difficulty understanding speech, inability to hear soft sounds, struggles in noisy environments, and feelings of social isolation. The etiology of deafness can be attributed to congenital factors such as genetic mutations or maternal infections, as well as acquired factors like exposure to loud noise, aging, medications, infections, head injuries, or diseases. Deafness can vary in severity and affect one or both ears. Proper diagnosis by healthcare professionals is essential for determining the specific type and cause of deafness, leading to appropriate treatment options such as hearing aids, cochlear implants, or assistive listening devices. [4]

3.1.5.Medical Causes:

Hearing loss can have various causes, including hereditary disorders where malformations of the inner ear are passed down through genes. Genetic disorders like osteogenesisimperfecta, Trisomy 13, and Treacher Collins syndrome can also contribute to hearing loss. Prenatal exposure to diseases such as rubella, influenza, and mumps can result in congenital deafness, as can exposure to substances like methyl mercury and certain medications. Noise-induced hearing loss can occur from prolonged exposure to loud noises, while trauma, diseases like meningitis and mumps, and conditions like Meniere's disease or exposure to certain chemicals can also lead to hearing loss.[5]

3.2. Muteness:

3.2.1. Introduction:

Muteness or Mutismis medically defined as a speech affliction wherein the patient is unable to have the normal capacity to speak resulting in the complete absence or at least a significant loss of verbal communication and is charted under both psychiatric and neurological diseases.

Rarely occurring as an isolated disorder, it is often prevalent in association with other ailments pertaining to cognitive abilities, disturbances in behavior or a related physiological disorder. Owing to its widespread existence and a host of causal factor triggering it, an effective understanding of this disability involves an in-depth study of the underlying neurological and psychological issues that form the basis of mutism. [6]

3.2.2. Historical Background:

The historical background of muteness reflects a complex tapestry of cultural beliefs, medical understanding, and societal attitudes. While ancient superstitions and prejudices have gradually given way to more enlightened perspectives, there is still progress to be made in creating a society that embraces and accommodates the diverse communication needs of individuals who are mute. [7]

3.2.3. Types of Muteness:

Muteness, the condition characterized by the inability to speak or limited speech ability, can manifest in various types based on its underlying causes. These include:

- Congenital muteness, which is present from birth and often linked to structural abnormalities or developmental issues.
- Selective mutism, where individuals can speak but consistently fail to do so in certain situations or around specific people.
- Neurogenic muteness, caused by neurological disorders or brain damage.
- Psychogenic muteness, arising from psychological or emotional factors.

- Functional muteness, a temporary or intermittent loss of speech without an apparent organic or psychological cause.

3.2.4.Symptomsand Etiology:

Muteness, characterized by the inability to speak or limited speech ability, can have various symptoms and underlying causes. Symptoms may include the absence of speech or difficulty in articulating sounds and words.

Etiology can be diverse, ranging from congenital conditions such as structural abnormalities in the vocal cords or brain, neurological disorders like apraxia or cerebral palsy, to psychological factors such as selective mutism or trauma. Other potential causes may involve damage to the vocal cords or larynx, hearing loss, or developmental delays.

Accurate diagnosis and evaluation by medical professionals are essential to determine the specific etiology of muteness and provide appropriate interventions and support for individuals affected by this condition.

3.2.5.Medical Causes:

"Mute Patients", Muteness from endotracheal intubation, tracheostomy, or damage to the vocal cords or trachea from disease or trauma can be extremely frustrating for patients.[8]

The situation can be equally frustrating for pharmacists, who rely on verbal information from patients when obtaining patient data and monitoring response to therapy. Written communication and the use of point-and-spell letter boards can be time consuming but often are the only means for two-way communication. Encourage these techniques and allow sufficient time for adequate communication. In addition, maintain your end of the conversation and do not limit your verbal responses just because the patient is mute.

4. Communication:

4.1. Introduction:

Communication, the exchange of meanings between individuals through a common system of symbols. [9]

Communication is a fundamental aspect of human interaction and plays a vital role in our daily lives. It serves as the foundation for sharing information, exchanging ideas, expressing emotions, and building relationships. Communication takes various forms, including verbal, nonverbal, written, and visual, and it occurs through different channels such as face-to-face conversations, written messages, digital platforms, and mass media.

Effective communication enables individuals to understand and be understood, fostering understanding, collaboration, and harmony in personal, social, and professional contexts. It is through communication that we convey our thoughts, transmit knowledge, negotiate, inspire, and connect with others on both intellectual and emotional levels. In essence, communication is the cornerstone of human interaction, enabling us to express ourselves, build connections, and navigate the complexities of the world we live in.

4.2. Types of communication:

Communication encompasses various types through which individuals exchange information, ideas, and emotions.

Verbal communication involves spoken or written words, while nonverbal communication relies on body language and gestures. Visual communication utilizes visual elements, and written communication involves written words. Interpersonal communication occurs in face-to-face interactions, while group communication involves exchanges within a team or group.

Mass communication reaches a wide audience through media channels, and digital communication takes place using digital platforms and technologies. These diverse forms of communication facilitate effective information sharing and connection in different contexts.

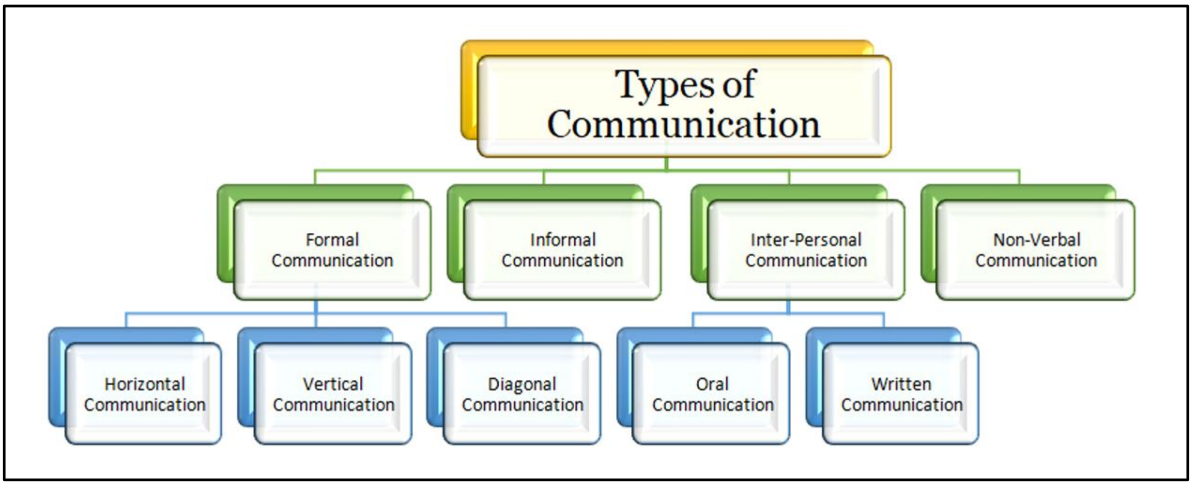

Fig 2. Types of communication.

5. Sign Language:

5.1. Introduction:

Sign language, any means of communication through bodily movements, especially of the hands and arms, used when spoken communication is impossible or not desirable. The practice is probably older than speech.

Sign language may be as coarsely expressed as mere grimaces, shrugs, or pointings; or it may employ a delicately nuanced combination of coded manual signals reinforced by facial expression and perhaps augmented by words spelled out in a manual alphabet.

Wherever vocal communication is impossible, as between speakers of mutually unintelligible languages or when one or more would-be communicators is deaf, sign language can be used to bridge the gap.[10]

5.2. Historical Background:

The first person credited with the creation of a formal sign language for the hearing impaired was Pedro Ponce de León , a 16th-century Spanish Benedictine monk. His idea to use sign language was not a completely new idea. Native Americans used hand gestures to communicate with other tribes and to facilitate trade with Europeans. Benedictine monks had used them to convey messages during their daily periods of silence. [11]

The recorded history of sign language in Western societies starts in the 17th century, as a visual language or method of communication, although references to forms of communication using hand gestures date back as far as 5th century BC Greece. Sign language is composed of a system of conventional gestures, mimic, hand signs and finger spelling, plus the use of hand positions to represent the letters of the alphabet. Signs can also represent complete ideas or phrases, not only individual words.

Most sign languages are natural languages, different in construction from oral languages used in proximity to them, and are employed mainly by deaf people in order to communicate. Many sign languages have developed independently throughout the world, and no first sign language can be identified. Both signed systems and manual alphabets have been found worldwide. Until the 19th century, most of what we know about historical sign languages is limited to the manual alphabets (fingerspelling systems) that were invented to facilitate transfer of words from an oral to a sign language, rather than documentation of the sign language itself.

5.3. Advantages of Sign Language:

- Its accessibility to individuals who are deaf or hard of hearing.
- Promoting language equality as a complete linguistic system.
- Enhancing communication through visual cues.
- Facilitating cultural expression within the Deaf community.
- Fostering multilingualism.

5.4. Disadvantages of Sign Language:

- Limited Understanding
- Interpreter Dependency
- Communication Barriers
- Linguistic Variations
- Technological Limitations

5.5. Different Sign Language Systems around the world:

Sign language is not a universal language but rather a diverse collection of languages that vary around the world. Here are some examples of different sign language systems:

- American Sign Language (ASL)
- British Sign Language (BSL)
- Australian Sign Language (Auslan)
- French Sign Language (FSL)

- Japanese Sign Language (JSL)
- Chinese Sign Language (CSL)
- South African Sign Language (SASL)

5.6. Some Characteristics of Sign Language:

Sign languages, as distinct visual-spatial languages, possess several unique characteristics that set them apart from spoken languages. Here are some key characteristics of sign language:

Visual-Gestural Modality: Sign language utilizes visual and gestural elements, relying on the hands, facial expressions, body movements, and space to convey meaning.

Iconicity: Sign languages often exhibit iconicity, meaning that signs can resemble the objects, actions, or concepts they represent.

Grammar and Syntax: Sign languages have their own grammatical rules and sentence structures. They use non-manual markers (facial expressions, body movements, and head tilting) to convey grammatical information such as tense, negation, and question formation.

Spatial Grammar: Sign languages make use of space to indicate spatial relationships, pronouns, and verb agreement.

Cultural Variation: Sign languages are influenced by and reflect the cultures and communities in which they are used.

Language Evolution: Like spoken languages, sign languages evolve and change over time. They adapt to new technologies, cultural shifts, and linguistic influences, allowing for the development of new signs and expressions.

Bilingualism and Multilingualism: Many deaf individuals and sign language users are bilingual or multilingual, using sign language alongside a spoken language.

5.7. Conclusion:

In conclusion, our exploration of the human ear, deafness, muteness, sign language communication, and sign language has provided valuable insights into the intricacies of human communication.

Understanding the ear's anatomy and functioning has revealed how we perceive sound and maintain balance. We have examined the different types, symptoms, and causes of deafness and muteness, recognizing their interconnectedness.

Sign language communication has been highlighted as a significant alternative for individuals with hearing or speech impairments, acknowledging both its advantages and limitations.

Additionally, our journey into sign language has deepened our appreciation for its linguistic richness and cultural importance. By fostering understanding in these areas, we promote inclusivity, empathy, and effective communication within our diverse society.

Chapter II: Artificial intelligence and Computer vision

Chapter II: Artificial intelligence and Computer vision

1. Introduction:

Computer Vision, often abbreviated as CV, is defind as a field of study that seeks to develop techniques to help computers "see" and understand the content of digital image and video. It is nothing but a scientific field that allows computers to capture, interpret, understand, and process the objects that are visually perceivable. With the help of Artificial Intelligence and deep learning models, computer vision systems are able to understand the captured digital images and react suitably. It is a multidisplinary field that could be called a sub field of artificial intelligence(AI) and machine learning (ML), it uses some specialized methods and make use of algorithms.[12]

Today, deep learning techniques are most commonly used for computer vision. This chapter explores the various ways deep learning is applied to computer vision. You will discover the benefits of using Convolutional Neural Networks (CNN), which provide a multi-layer architecture that allows neural networks to focus on the most relevant features of an image.

The performance and efficiency of a CNN are determined by its architecture. This includes the structure of the layers, how the elements are designed, and the elements present in each layer. Many CNNs have been created, but the following models in this chapter are among the foundational models for computer vision.

2. Computer Vision:

Computer Vision is a field of machine learning dedicated to interpreting and understanding images and videos. It is used to teach computers to "see" and utilize visual information to perform visual tasks, much like humans do.

Computer vision models are designed to translate visual data based on identified features and contextual information learned during training. This enables models to interpret images and videos and apply these interpretations to prediction or decision-making tasks.

Although both related to visual data, image processing is not the same as computer vision. Image processing involves manipulating or enhancing images to produce a new result. This can include optimizing brightness or contrast, increasing resolution, blurring sensitive information, or cropping. The difference between image processing and computer vision is that the former does not necessarily require the identification of content. [13]

3. Artificial Intelligence (AI):

Artificial Intelligence is the general term that encompasses both machine learning and deep learning. As you can see in the diagram, even deep learning is a subset of machine learning. So all three, AI, machine learning, and deep learning, are just subsets of each other. Now let's dive in and understand how they are exactly different from one another.

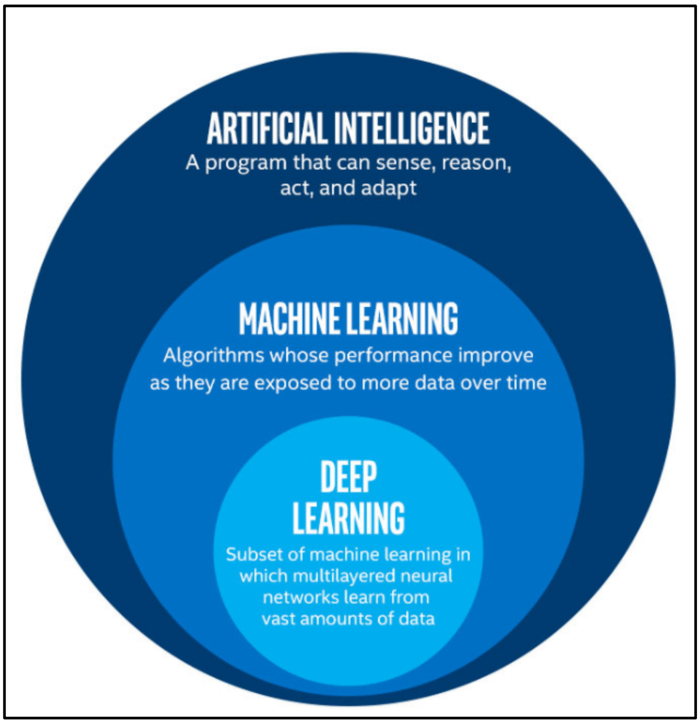

Fig 3. The relationship between AI and Ml and DL

3.1. Definition:

Artificial Intelligence (AI) is a field of computer science that aims to create intelligent machines capable of mimicking human intelligence. It involves technologies like machine learning and deep learning, enabling systems to process data, learn from patterns, and make informed decisions. AI can be categorized into Narrow AI, specialized for specific tasks, and General AI, which strives to achieve human-level intelligence. The applications of AI are vast, spanning industries such as healthcare, finance, transportation, and entertainment. However, ethical concerns and responsible guidelines are essential to address potential risks and ensure the responsible use of AI.

3.2. Machine Learning:

Machine learning refers to computer programs that can learn on their own without explicit human programming. The term originated from Alan Turing's 1950 paper "Computing Machinery and Intelligence," which introduced the concept of a "learning machine" capable of convincing humans of its authenticity. Today, machine learning encompasses various programs used in big data analysis and exploration. Machine learning algorithms power predictive programs like spam filters, product recommendations, and fraud detectors. Understanding supervised and unsupervised machine learning, ensemble modeling, and semi-supervised learning is essential for data scientists.

Supervised learning involves training programs to generate responses based on known and labeled datasets. Classification and regression algorithms, such as decision trees and support vector machines, are commonly used in supervised learning. On the other hand, unsupervised machine learning uses algorithms to generate responses on unknown and unlabeled data. Data specialists employ unsupervised techniques to uncover patterns in new datasets, often utilizing clustering algorithms like K-means.

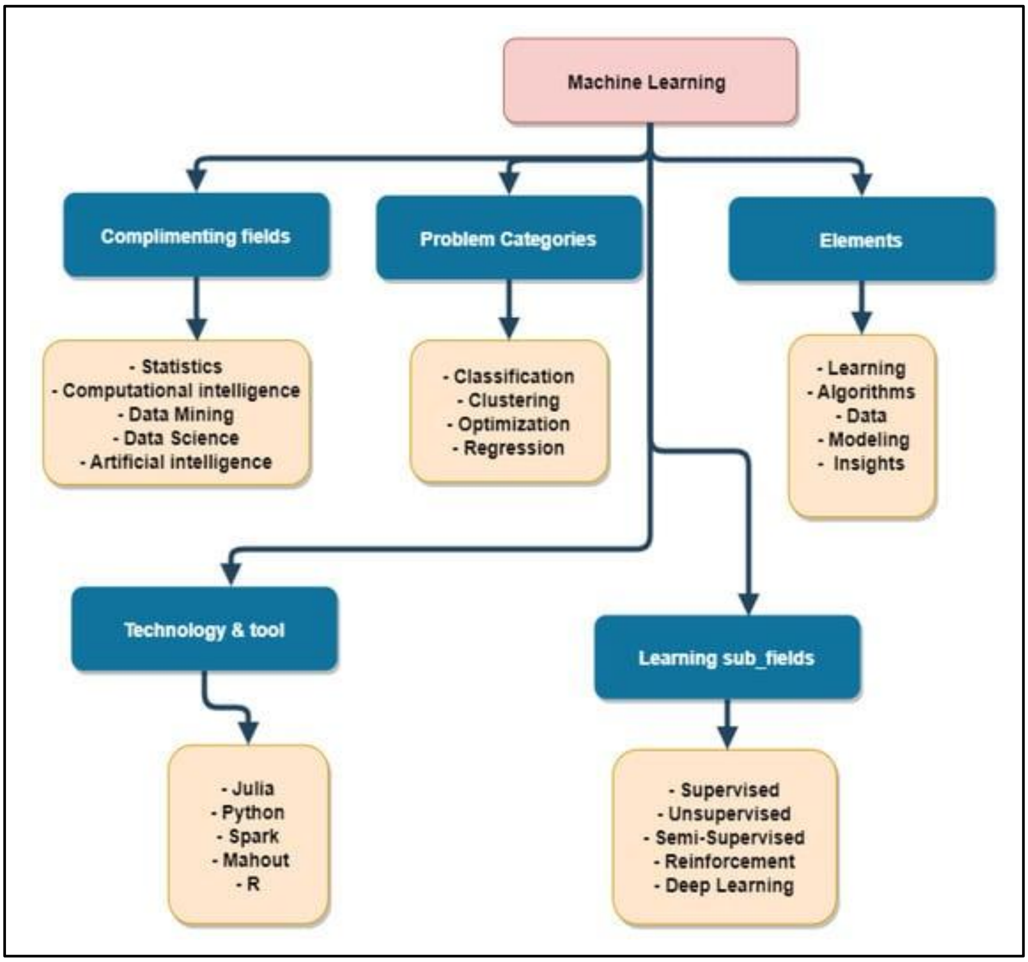

Fig 4. Machine Learning Parts.

3.3. Difference between Machine Learning (ML) and Deep Learning (DL):

Machine learning (ML) works effectively when we have a moderate amount of labeled data that we train and obtain desired results based on the problem we need to solve. The functioning of machine learning does not depend on the type of system or hardware we use; it works efficiently on all systems. Machine learning takes less time to train the data and be ready to analyze results. It works by reducing the complexity of the data to a lower level so that result analysis is more efficient and produces more accurate outcomes.

Deep Learning (DL) is also used to implement artificial intelligence techniques through neural networks. The basic principle is the neural network on which deep learning operates. Deep learning takes more time to train the data but produces more accurate results compared to machine learning. Deep learning requires advanced machines to operate efficiently and yields more precise results when a large amount of data is used to analyze the outcomes through deep learning.

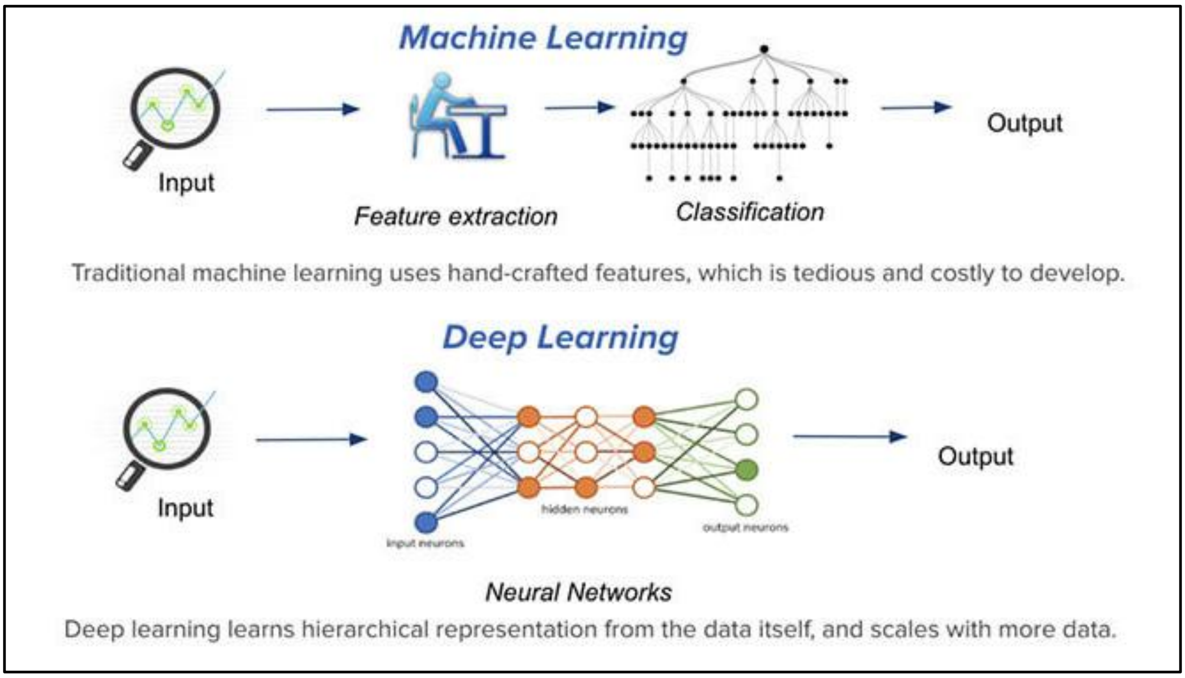

Fig 5. Difference between ML and DL.

4. Classification of Images:

"Image classification is a key step in computer vision as it allows us to identify what an image contains. Visual analysis of data has now become an essential component of data analysis. Many aspects of our daily lives have been impacted by recent technological advancements, particularly in the field of artificial intelligence. Medical technologies that enhance the detection of certain conditions, as well as smart devices, have made our lives easier. However, these revolutionary advancements are often built upon complex mechanisms. In the following, we will define and discuss the principles of image classification." [14]

4.1. Principle of operation in general:

In general, the computer analyzes an image in the form of pixels. It does so by considering the image as an array of matrices, with the size depending on the image resolution. To simplify, image classification from the computer's perspective involves analyzing these statistical data using algorithms. In digital image processing, image classification is performed by automatically grouping pixels into specific categories known as 'classes'.

5. Classification Methods:

5.1. Supervised Classification:

- We have pre-labeled elements.

Example: Articles categorized as economy, politics, sports, culture, etc.

- We want to classify a new element.

Example: Assigning a label such as economy, politics, sports, etc., to it.

5.2. Unsupervised Classification:

-We have unlabeled elements.

Example: Words in a text.

-We want to group them into classes.

Example: If two words have the same label, they are related to the same theme.

6. Presentation of certain classification techniques:

6.1. K-Nearest Neighbors (K-NN) Classifier:

The K-Nearest Neighbors (K-NN) algorithm is a supervised learning method that can be used for regression and classification tasks. Its operation can be likened to "tell me who your neighbors are, and I'll tell you who you are".

Unlike logistic regression, K-NN does not compute a predictive model based on the training dataset to make predictions. In fact, K-NN does not need to construct a predictive model at all. Therefore, K-NN does not have a dedicated learning phase. This is why it is sometimes referred to as a lazy learning algorithm.

To make predictions, K-NN relies on a dataset to produce results. [15]

6.2. Decision Tree:

A decision tree is a supervised learning algorithm that is well suited for classification problems as it can categorize classes at a granular level. It operates like a flowchart, dividing data points into two similar categories at each step, starting from the "trunk" of the tree to the "branches" and finally to the "leaves", where the categories become more finely similar. This creates nested categories, enabling organic classification with limited human supervision.

6.3. Support Vector Machine (SVM):

Support Vector Machines (SVMs) are a class of algorithms based on the principle of minimizing the "Structural Risk" described using statistical learning theory - linear separation. This involves separating individuals represented by the hyperplane in a space with dimensions equal to the number of features, thereby dividing the individuals into two classes. When the data to be classified is linearly separable, this separation can be achieved. Otherwise, the data is projected into high-dimensional spaces where they become linearly separable. [16]

Here are some applications of SVM:

- Text and hypertext classification
- Image classification
- Handwritten character recognition
- Biological sciences, including protein classification

SVMs are popular due to their ability to handle high-dimensional data and perform well in both linear and non-linear classification tasks. They find an optimal hyperplane that maximally separates data points of different classes, considering a wide margin. Support vectors, the closest data points to the decision boundary, are instrumental in defining the hyperplane. SVMs excel at handling complex datasets and effectively managing outliers. They can also utilize various kernel functions to handle non-linear decision boundaries.

7. Artificial neural network (ANN):

7.1. Definition:

An artificial neural network (ANN) is a computational model inspired by the structure and functioning of biological neural networks in the human brain. It is a machine learning algorithm designed to recognize patterns, learn from data, and make predictions or decisions. ANN consists of interconnected nodes, called artificial neurons or nodes, organized into layers.

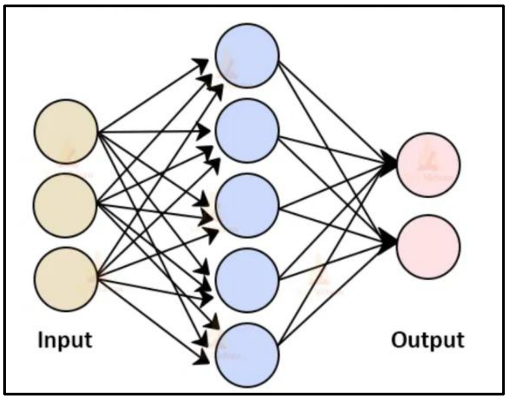

Fig 6. Artificial neural network (ANN)

7.2. Biological Basis of Artificial Neural Networks:

Artificial neural networks are a technology based on studies of the brain and nervous system as depicted in Fig.7. These networks emulate a biological neural network but they use a reduced set of concepts from biological neural systems. Specifically, ANN models simulate the electrical activity of the brain and nervous system. Processing elements (also known as either a neurode or perceptron) are connected to other processing elements. Typically the neurodes are arranged in a layer or vector, with the output of one layer serving as the input to the next layer and possibly other layers. A neurode may be connected to all or a subset of the neurodes in the subsequent layer, with these connections simulating the synaptic connections of the brain. Weighted data signals entering a neurode simulate the electrical excitation of a nerve cell and consequently the transference of information within the network or brain. The input values to a processing element, in, are multiplied by a connection weight, wn,m, that simulates the strengthening of neural pathways in the brain. It is through the adjustment of the connection strengths or weights that learning is emulated in ANNs.[17]

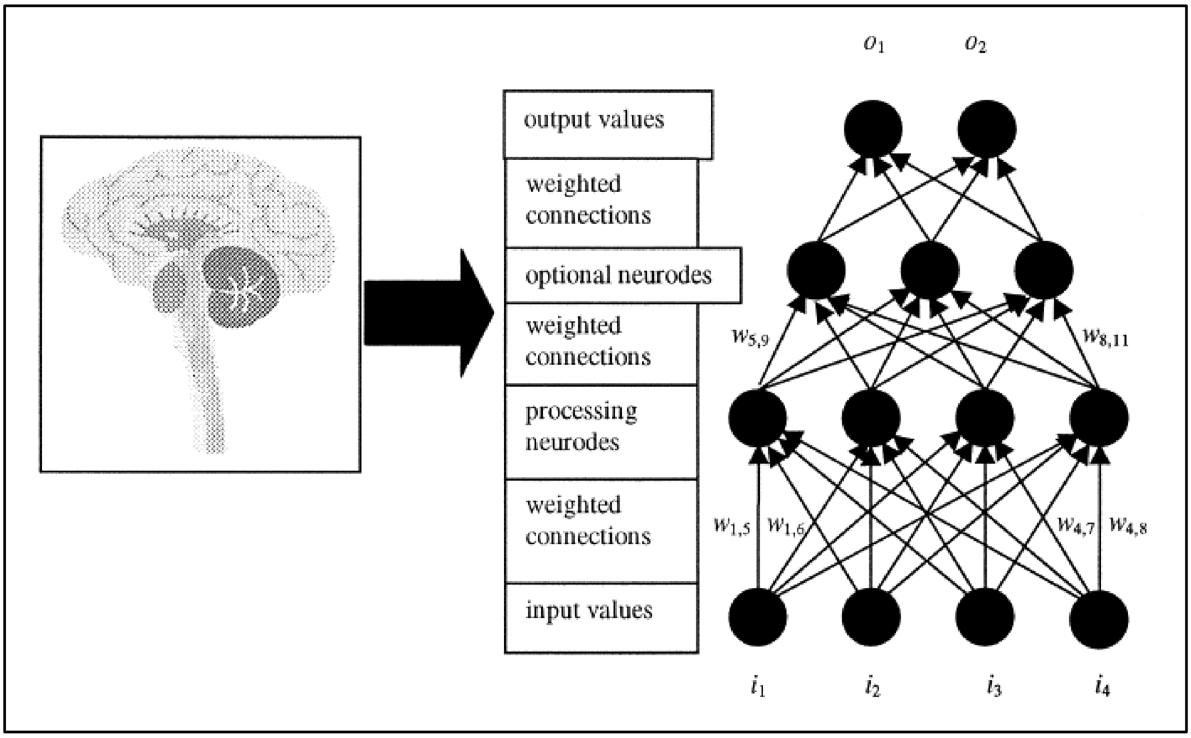

Fig 7. Biological Basis of Artificial neural network (ANN).

7.3. The architecture of an artificial neuron:

An artificial neuron, also known as a perceptron, is the basic building block of an artificial neural network. It takes multiple inputs, each multiplied by a corresponding weight, and produces an output by applying an activation function.

An artificial neuron is an artificial and schematic representation of a biological neuron:

- Synapses are modeled by weights.
- The soma or cell body is modeled by the transfer function, also called function activation.
- The axon through the exit element.

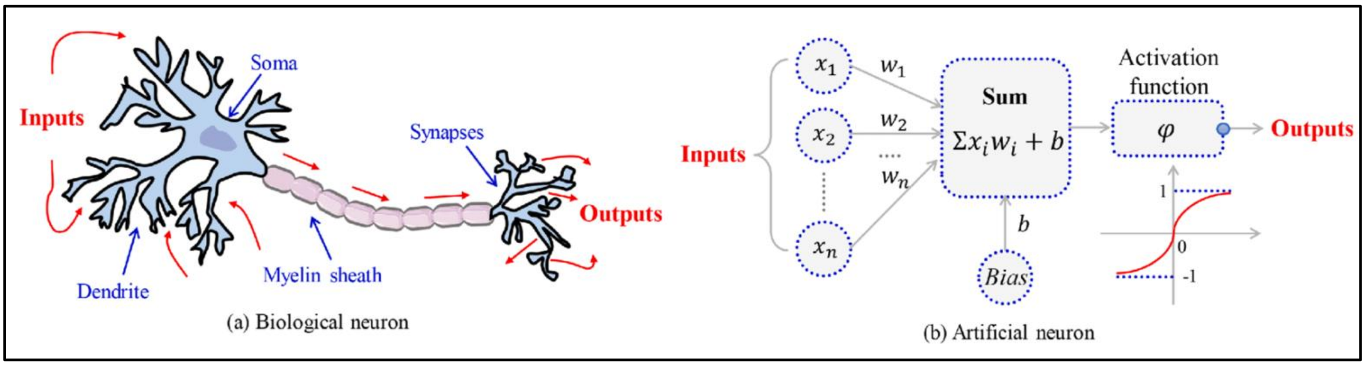

Fig 8. Comparison between biological neuron and artificial neuron.

7.4. Operation of an artificial neural network:

The concept of neural networks is based on three main steps:

1. Weight the input value for each neuron in the layer.
2. Then, for each layer, the weights of all neurons are summed and add bias.
3. Finally, apply the activation function to this value to calculate the new output.

8. ANN typology:

8.1. Perceptron:

The Perceptron was invented in 1957 by Frank Rosenblatt at the Cornell Aeronautics Laboratory. Based on the first concepts of artificial neurons, he proposed the “Perceptron learning rule“.

A Perceptron is an artificial neuron, and thus a neural network unit. It performs computations to detect features or patterns in the input data. It is an algorithm for supervised learning of binary classifiers. It is this algorithm that allows artificial neurons to learn and process features in a data set. [18]

8.1.1. The function of the Perceptron:

In reality, the Perceptron is a mathematical function. The input data (x) is multiplied by the weight coefficients (w). The result is a value.

This value can be positive or negative. The artificial neuron is activated if the value is positive. It is therefore activated only if the calculated weight of the input data exceeds a certain threshold.

The predicted result is compared with the known result. If there is a difference, the error is back propagated in order to adjust the weights.

The perceptron is a supervised learning algorithm primarily used as a binary classifier, dividing data into two categories. It accomplishes this by creating a hyperplane that separates the input space into the respective categories.

Advantages of Perceptron:

- Logical Gate Implementation: Perceptrons are capable of implementing logic gates such as AND, OR, and NAND, making them useful for basic logical operations.

Disdvantages of Perceptron:

- Limited to Linearly Separable Problems: Perceptrons can only learn and classify linearly separable problems, where a single straight line or hyperplane can effectively divide the data into categories. They struggle with nonlinear problems, such as the Boolean XOR problem, where the data cannot be separated by a single straight line or hyperplane.

8.1.2. Multilayer Perceptron (MLP):

The Multilayer Perceptron (MLP) is a neural network that consists of multiple layers of artificial neurons. It utilizes forward and backward propagation to process input data. Forward propagation involves multiplying inputs by weights and passing them through activation functions. Backpropagation adjusts the weights to minimize loss. MLPs have an input layer, output layer, and multiple hidden layers. Activation functions, often nonlinear, are applied, with softmax commonly used in the output layer. MLPs serve as a gateway to complex neural networks.

- Applications of Multilayer Perceptron:
  - Speech Recognition: MLPs have been applied to speech recognition tasks, such as converting spoken words into written text.
  - Automatic Translation: MLPs have been utilized in automatic translation systems, aiding in translating text or speech from one language to another.
  - Complex Classification: MLPs are effective in complex classification tasks, such as classifying images into multiple categories.
- Advantages of Multilayer Perceptron:

  - Suitable for Deep Learning:
  MLPs are used in deep learning due to their fully connected layers and the availability of back propagation for weight adjustments.

- Disadvantages of Multilayer Perceptron:

  - Complex to Design:

  MLPs can be challenging to design and configure due to the complex structure and the need to determine appropriate activation functions, layer sizes, and hyper parameters.

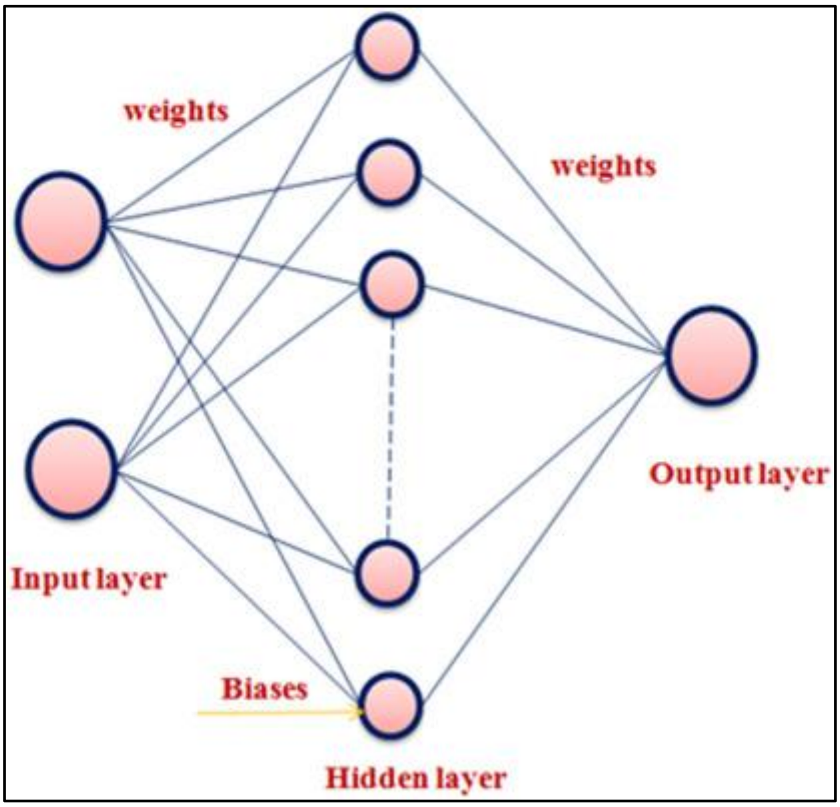

Fig 9. MLP neural network.

8.2. Recurrent Neural Network (RNN):

A Recurrent Neural Network (RNN) is a type of artificial neural network that is designed to process sequential data by incorporating feedback connections. Unlike feedforward neural networks, which process data in a strictly forward direction, RNNs have recurrent connections that allow information to be carried across previous time steps, enabling them to capture temporal dependencies in the data.

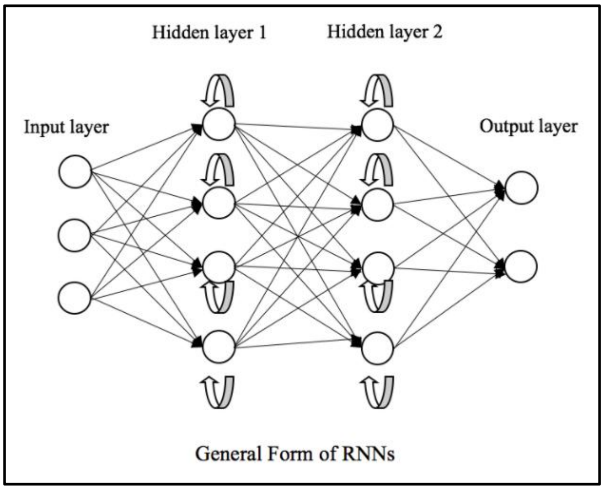

Fig 10. Recurrent Neural Network.

The architecture of an Recurrent Neural Network :

The architecture of an RNN consists of repeating units, typically referred to as cells, which are interconnected in a recurrent manner. Each cell takes an input, which can be the current data point in a sequence, along with the previous hidden state. The input and hidden state are processed through activation functions to produce an output and update the hidden state for the next time step. This process is repeated for each time step in the sequence, allowing the network to capture and utilize information from past inputs.

RNNs can be trained using the backpropagation through time (BPTT) algorithm, which extends the concept of backpropagation to handle the temporal nature of the network. During training, the weights and biases of the RNN are adjusted based on the error propagated back through time, optimizing the network's performance on the given task.

Applications of recurrent neural networks:

- Word processing like auto-suggest, text check, grammar, etc.
- Speech processing from text
- Image tagger
- Sentiment analysis
- Translation

Benefits of Recurrent Neural Networks:

Recurrent Neural Networks are powerful models for processing sequential data, allowing for the incorporation of time dependencies and providing valuable insights in various domains.

The defining characteristic of an RNN is its ability to maintain a hidden state or memory that persists and evolves as new inputs are processed. This hidden state serves as a form of short-term memory, allowing the network to remember and consider information from previous time steps when making predictions or decisions.

Drawbacks of Recurrent Neural Networks:

One common issue with traditional RNNs is the vanishing gradient problem, which can hinder the learning of long-term dependencies. To address this, variations of RNNs, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been developed with specialized memory cells that alleviate the vanishing gradient problem and better capture long-term dependencies.

9. Convolutional Neural Networks (CNN):

Convolutional Neural Networks (CNNs) have demonstrated excellent performance in numerous computer vision and machine learning problems. CNNs belong to the category of deep neural networks, which are primarily deployed in the field of image analysis and recognition. CNNs are particularly useful in various image-related tasks such as image classification, semantic segmentation, object detection, and more.

CNNs consist of multiple layers of artificial neurons. These artificial neurons, similar to the neural cells used by the human brain to transmit sensory input and other responses, are mathematical functions used to compute the sum of various inputs and provide an output in the form of an activation value.

When we feed an input image to a CNN, each of its internal layers generates different activation maps. These activation maps indicate relevant features of the given input image. Each of the CNN neurons typically takes input in the form of a group/patch of pixels, multiplies their values (colors) by its weight value, adds them up, and passes them through the respective activation function.

The first layer (or possibly the lower layer) of the CNN usually recognizes different features of the input image, such as horizontal, vertical, and diagonal edges.

The output of the first layer is used as input for the next layer, which in turn extracts other complex features from the input image, such as corners and combinations of edges. As we go deeper into the convolutional neural network, the layers start to detect higher-level features like objects, faces, and more. [19]

In summary, CNNs have proven to be powerful tools in computer vision and machine learning. They operate by extracting features from input images through multiple layers of artificial neurons. These networks have been successfully applied to various tasks, including image classification, object detection, and semantic segmentation. The hierarchical nature of CNNs allows them to capture increasingly complex features as we progress deeper into the network.

9.1. Comparison between CNN and RNN:

- CNNs excel in interpreting visual data and non-sequential data, while RNNs are designed for recognizing sequential or time-dependent data.

- CNNs are well-suited for applications like face detection, medical analysis, drug discovery, and image analysis. RNNs are valuable for language translation, entity extraction, conversational AI, and speech analysis.

- CNNs process fixed-size inputs and generate fixed-size outputs, whereas RNNs can handle inputs and outputs of variable lengths.

- CNNs are inspired by the connectivity pattern in the animal visual cortex, where neurons respond to overlapping regions in the visual field. RNNs leverage temporal information in time series data, incorporating the influence of past inputs on future outputs.[20]

9.2. CNN Architecture:

A CNN typically consists of five layers: a convolutional layer, a pooling layer, and a fully connected layer (FC).

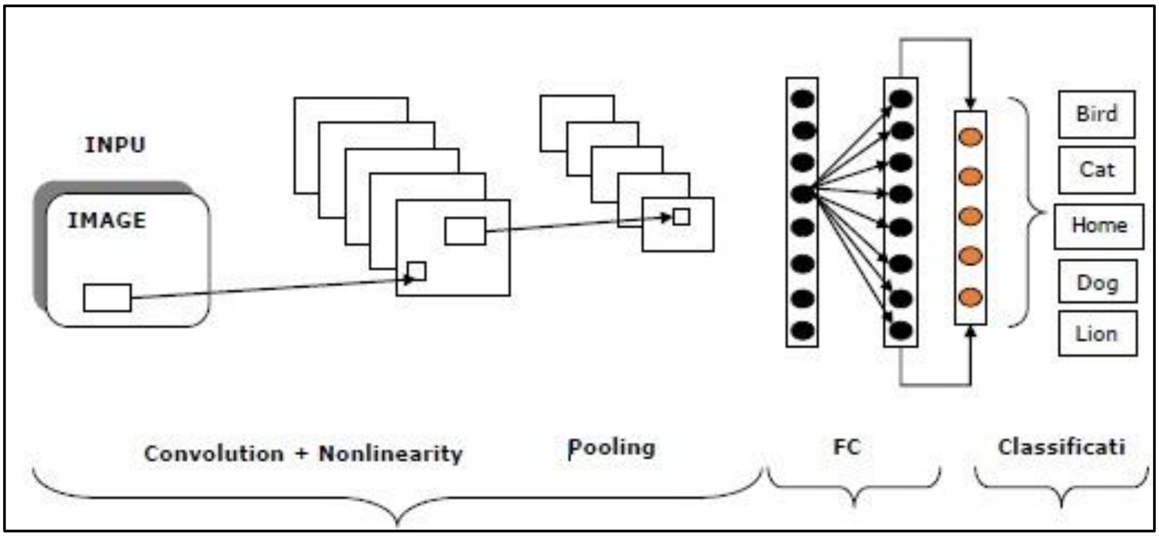

Fig 11. Architecture of CNN.

9.2.1.Input Image:

Pixels serve as the fundamental components of digital images, representing binary-encoded visual data arranged in a matrix form. Each pixel's value determines its brightness and color within the image.

When humans perceive an image, their brain rapidly processes vast amounts of information. Neurons in the brain have their own receptive fields and interconnectedness to cover the entire visual field. Similarly, in a CNN, each neuron analyzes data within its receptive field. CNN layers are designed to detect simple patterns like lines and curves, progressively advancing to more complex patterns such as faces and objects. Thus, CNNs have the potential to enable computers to gain visual capabilities.[21]

9.2.2. Convolutional Layer:

A convolutional layer is a fundamental component of CNN architecture that performs feature extraction. It typically involves a combination of linear and non-linear operations, namely convolution and activation functions.

Convolution:

Convolution, in mathematics, is an operation that combines two functions to produce a third function that expresses how the shape of one function is modified by the other.

Kernel Convolution:

A kernel (also called a filter) is a small 2D matrix whose content is based on the operations to be performed. A kernel maps the input image through matrix multiplication and addition. The resulting output is of lower dimensionality, making it easier to work with.[22]

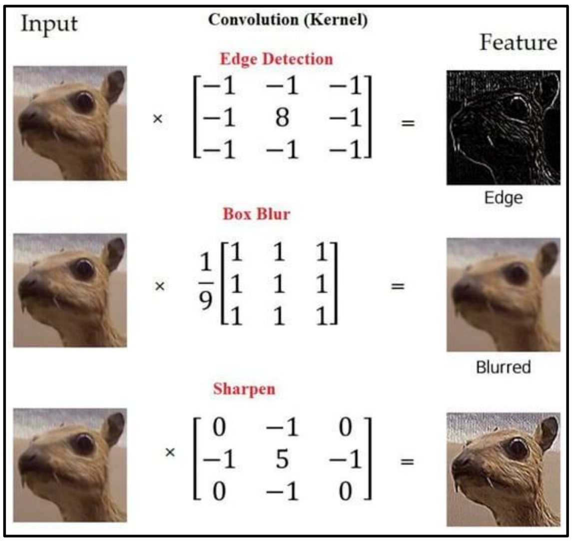

Fig 12. Effects of different convolution matrices.

The shape of the filter is heavily influenced by the shape of the input image and the overall architecture of the network. Typically, the kernels used in the convolutional layers have a size of MxM, representing a square matrix. The kernel's movement, or convolution, is performed by scanning from left to right and from top to bottom within the input image.

Stride:

Defines by which step the filter moves, for example:

- A stride of 1 moves the filter one row/column at a time.
- A stride of 2 moves the filter two rows/columns at a time.

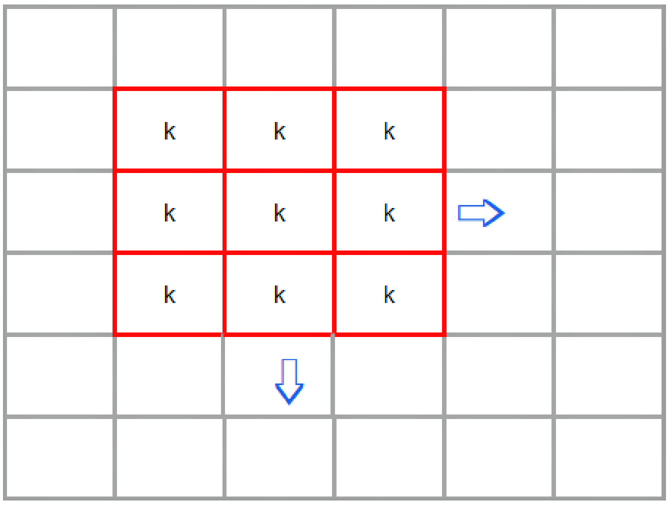

Fig 13. Working principle of stride.

Principle of operation of the convolution filter:

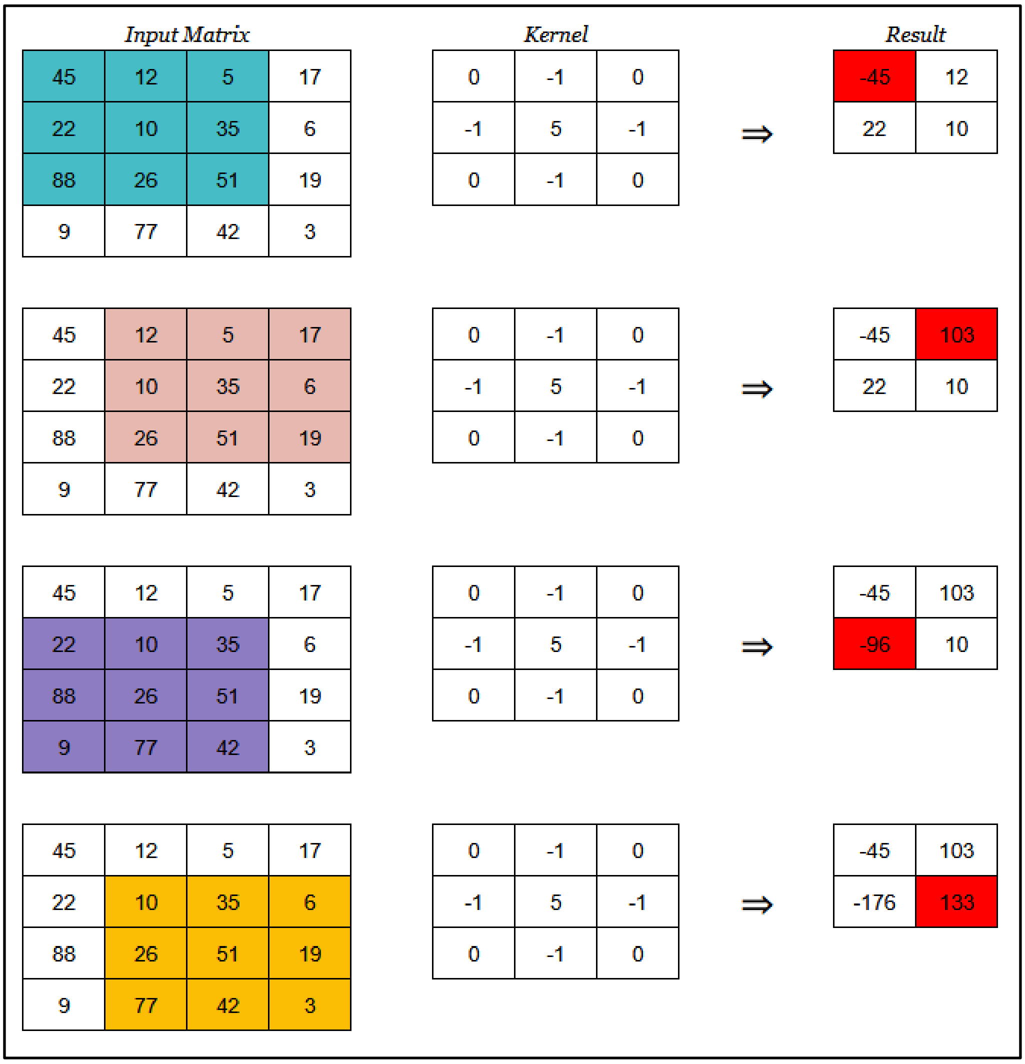

Fig 14. Operation of the convolution filter.

The first entry in the convolved result is calculated as follows:

45*0 + 12*(-1) + 5*0 + 22*(-1) + 10*5 + 35*(-1) + 88*0 + 26*(-1) + 51*0 = -45

The Secondentry in the convolved result is calculated as follows:

12*0 + 5*(-1) + 17*0 + 10*(-1) + 35*5 + 6*(-1) + 26*0 + 51*(-1) + 19*0 = 103

The Thirdentry in the convolved result is calculated as follows:

22*0 + 10*(-1) + 35*0 + 88*(-1) + 26*5 + 51*(-1) + 9*0 + 77*(-1) + 42*0 = -96

Thefourthentry in the convolved result is calculated as follows:

10*0 + 35*(-1) + 6*0 + 26*(-1) + 51*5 + 19*(-1) + 77*0 + 42*(-1) + 3*0 = 133

Padding :

Padding is a technique used in convolutional neural networks (CNNs) to preserve the spatial dimensions of the input volume or image when applying convolutional operations. It involves adding extra pixels or values around the border of the input before performing convolutions.

There are two commonly used types of padding in convolutional neural networks:

Valid Padding: No extra pixels are added around the input, resulting in a smaller output size as the filter cannot extend beyond the borders.

Same Padding: The output size of the convolutional layer is maintained equal to the input size by adding the necessary number of zero-filled pixels around the input. This helps preserve spatial information and prevents the loss of information during convolutional operations.

Activation function:

Activation functions are essential in neural networks, providing non-linearity to enable learning complex patterns. They determine the output of each neuron based on weighted inputs, allowing the network to model non-linear relationships.

Common activation functions include sigmoid, ReLU, Leaky ReLU, softmax, and tanh. Sigmoid maps values between 0 and 1, ReLU sets negative inputs to 0, Leaky ReLU introduces a small negative slope, softmax produces class probabilities, and tanh maps values between -1 and 1.

These functions enable neural networks to solve complex tasks by introducing non-linear transformations to the input data.

9.2.3. POOLING Layer:

Pooling is a feature in convolutional neural networks that involves reducing the dimensionality of the input data.

The two main approaches for pooling are average pooling, where values within a region are averaged, and max pooling, where the highest value is selected. Pooling is a downsampling step that helps to decrease computation time. Average pooling is advantageous for detecting weak signals like in steganalysis, while max pooling is effective for detecting strong signals such as objects.

Overall, pooling plays a key role in dimensionality reduction and improving computational efficiency in CNNs.[23]

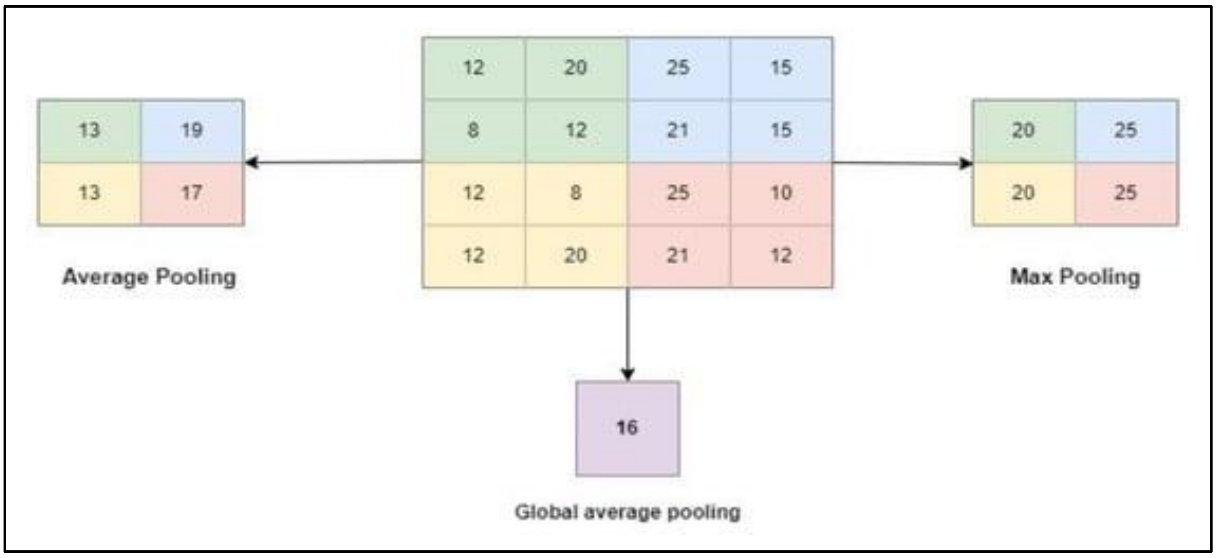

Fig 15. Pooling layer.

9.2.4. Fully connected layer:

The fully connected layer is a type of layer in a multilayer perceptron (MLP) consisting of three types of layers: input, hidden, and output.

- The input layer receives the features generated by the CNN.

- The hidden layer is a sequence of neurons whose weights are learned during the training phase. An MLP can have one or more hidden layers.

- The output layer is also a sequence of neurons, but it has a different activation function. Typically, the softmax function is used to generate probabilities for each category in the problem.[24]

9.2.5. Loss layer:

The loss layer, typically the last layer of the network, is responsible for calculating the loss function. Different loss functions can be applied depending on the specific task at hand.[25]

- For single-class prediction among K mutually exclusive classes, the "Softmax" loss function is commonly used.
- The sigmoid cross-entropy loss is utilized when predicting K independent probability values within the range of [0, 1].
- When performing regression tasks towards real values, the Euclidean loss (also known as Mean Squared Error) is commonly used.

9.3. Training of a CNN:

The initial step involves determining the network design, including the number of layers, their size, and the matrix operations that connect them. The network coefficients are then optimized to reduce the classification error during training.

The algorithm used to train a CNN is called back propagation, which utilizes the output value from the last layer to measure an error. This error value is used to update the weights of each neuron in that layer.

The updated weights are used to measure an error value and update the weights of the previous layers.

The algorithm repeats this process until it reaches the first layer.

For top-performing CNNs, this training process can take several weeks, with numerous GPUs working on hundreds of thousands of annotated images.

10. CNN architectures:

CNN topologies play a crucial role in determining the performance of a CNN. In recent years, numerous new network topologies have been proposed to improve accuracy or reduce computational complexity. A clear trend is that CNN networks are becoming deeper and more complex.

Today, we have access to larger datasets with millions of high-resolution labeled data points across thousands of categories, such as Image Net and Label Me. With the advent of powerful GPU machines, CNNs deliver exceptional performance in image classification tasks. In 2012, a deep convolutional neural network called Alex Net, designed by Krizhevsky, achieved remarkable results in the Image Net Large Scale Visual Recognition Challenge (ILSVRC). The success of Alex Net served as inspiration for various CNN models such as ZFNet, VGGNet, Google Net, ResNet, Dense Net, Caps Net, SENet, and more in the following years.

11. Examples of CNN architectures:

11.1. VGGNet (2014):

VGGNet, introduced in 2014, is a CNN architecture developed by Karen Simonyan, Andrew Zisserman, and their team at the University of Oxford. It consists of 16 convolutional layers and is similar to AlexNet, utilizing 3x3 convolutions with a large number of filters. VGGNet was trained for several weeks on multiple GPUs and has become a popular choice for image feature extraction. One key innovation of VGGNet is the stacking of multiple convolutional layers with smaller filters, reducing the number of features. Additionally, the inclusion of 3 ReLU layers enhances the network's learning capacity. VGGNet has made significant contributions to the field of CNNs and has been influential in various computer vision tasks, particularly in extracting features from images. [26]

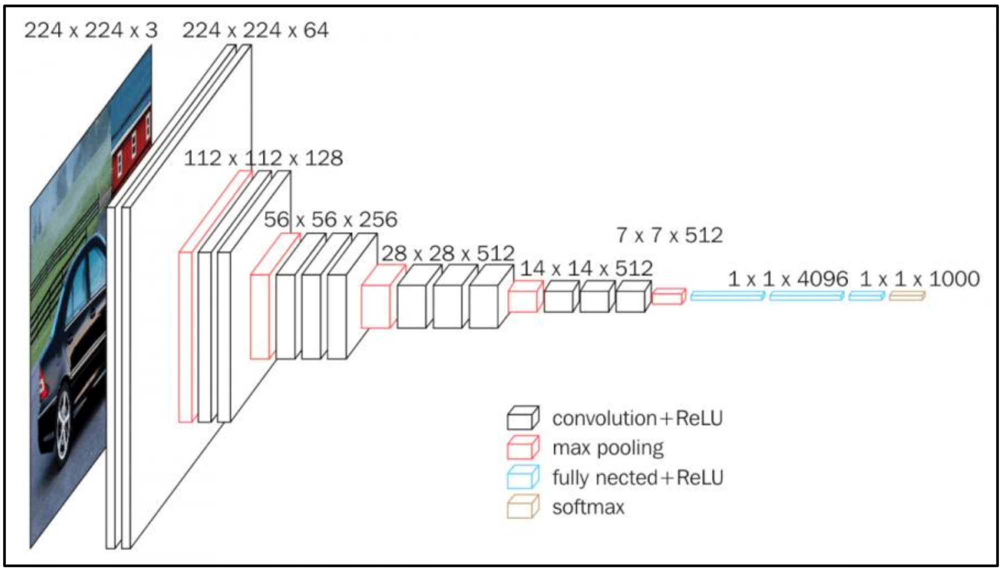

Fig 16. VGGNet Architecture.

- **Advantages:**
  - VGGNet brought a significant improvement in accuracy and speed, mainly due to its increased model depth and the introduction of pre-trained models.
  - The increase in the number of layers with smaller kernels resulted in increased non-linearity, which is beneficial in deep learning.

- VGGNet introduced various architectures built on a similar concept, providing more options for choosing the most suitable architecture for specific applications.

▪ **Disadvantages:**

- VGGNet is slower compared to the newer ResNet architecture, which introduced the concept of residual learning, representing another major advancement.

11.2. ResNet(2015):

ResNet, short for Residual Network,is a deep learning model designed for computer vision tasks. It addresses the vanishing gradient problem by introducing skip connections, allowing information to bypass certain layers. This accelerates training and enables the construction of very deep networks that can capture complex patterns in visual data effectively.[27]

Residual blocks are an important part of the ResNet architecture. In older architectures such as VGG16, convolutional layers are stacked with batch normalization and nonlinear activation layers such as ReLu between them. This method works with a small number of convolutional layers—the maximum for VGG models is around 19 layers. However, subsequent research discovered that increasing the number of layers could significantly improve CNN performance.

The ResNet architecture introduces the simple concept of adding an intermediate input to the output of a series of convolution blocks. This is illustrated below.

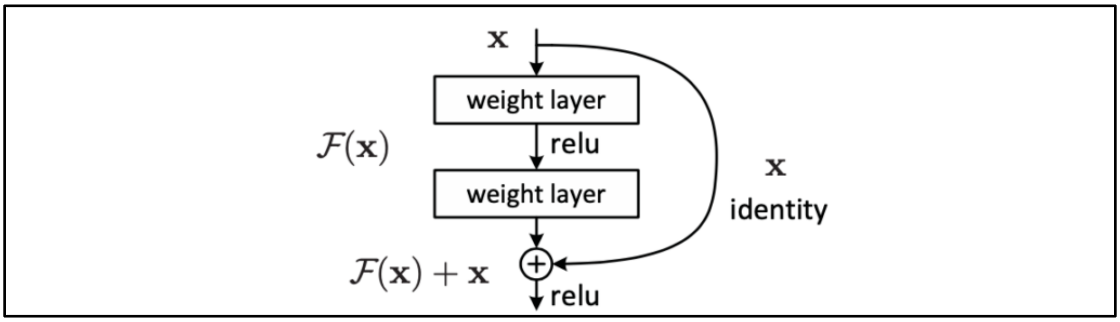

Fig 17. Typical residual block.

11.2.1. Comparison between Advantages and Disadvantages in ResNet:

ResNet, the revolutionary deep learning model for computer vision, offers significant advantages over previous architectures. It improves convergence, achieves higher accuracy, and enables the learning of intricate representations. However, ResNet also has drawbacks, including increased complexity, potential for overfitting, higher memory requirements, reduced interpretability, and the need for domain-specific tuning.

11.3. Xception (2016):

Xception is a deep convolutional neural network (CNN) architecture that was introduced in 2016. It was designed to improve the efficiency and accuracy of image classification tasks.

Xception is a deep convolutional neural network architecture that involves Depth wise Separable Convolutions. It was developed by Google researchers. Google presented an interpretation of Inception modules in convolutional neural networks as being an intermediate step in-between regular convolution and the depth wise separable convolution operation (a depth wise convolution followed by a point wise convolution). In this light, a depth wise separable convolution can be understood as an Inception module with a maximally large number of towers. This observation leads them to propose a novel deep convolutional neural network architecture inspired by Inception, where Inception modules have been replaced with depth wise separable convolutions.[28]

The Xception architecture:

The data first goes through the entry flow, then through the middle flow, which is repeated eight times, and finally through the exit flow. Note that all Convolution and Separable Convolution layers are followed by batch normalization.

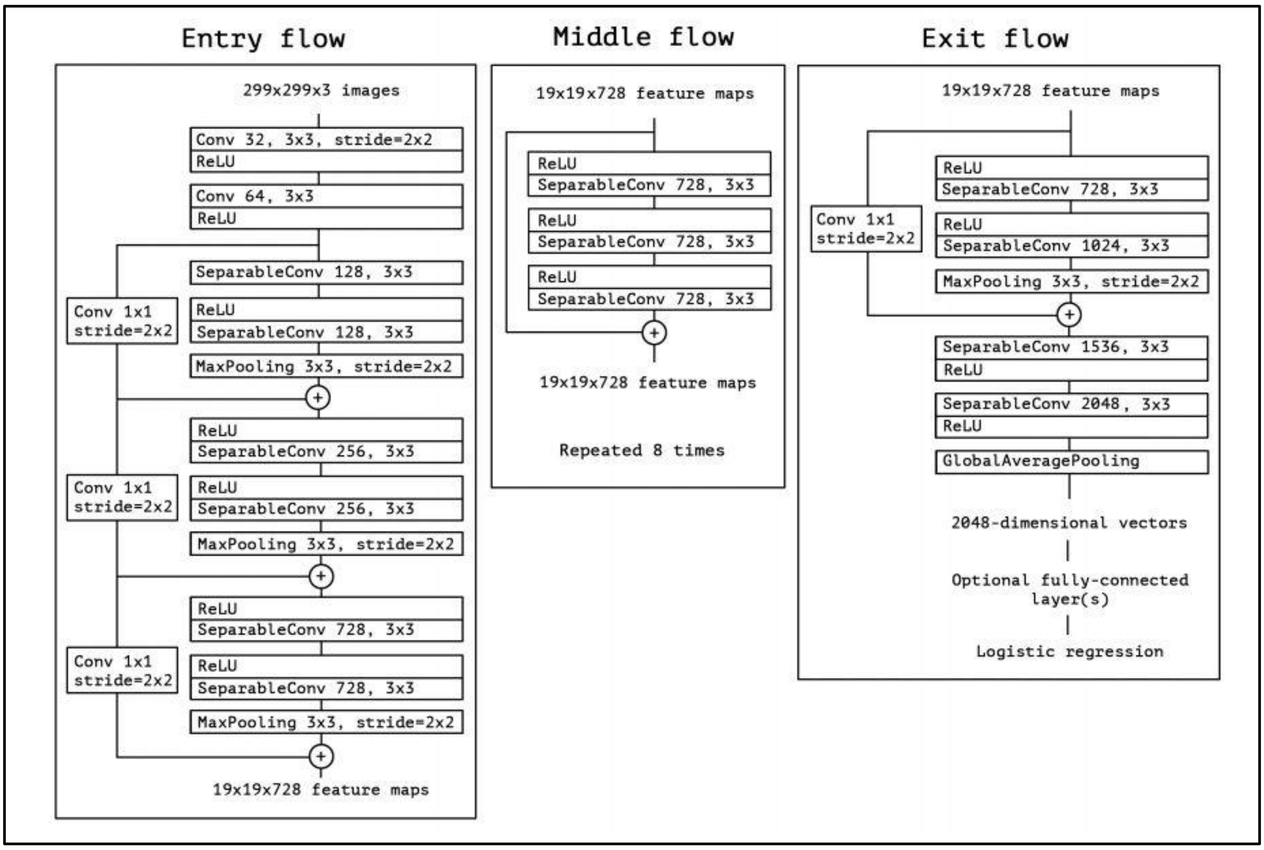

Fig 18. Xception architecture.

Advantages of Xception:

1. Improved performance in image classification tasks with effective capture of fine-grained details and spatial relationships.
2. Parameter efficiency, resulting in smaller model size and faster training and inference times.
3. Computational efficiency, making it suitable for real-time applications with low latency requirements.
4. Ability to learn hierarchical features, capturing both low-level and high-level information in complex patterns.

Disadvantages of Xception:

1. Limited contextual understanding, particularly in tasks requiring long-range dependencies or global context.

2. Sensitivity to noisy data during training, necessitating careful preprocessing and regularization.
3. Increased training complexity, requiring more computational resources and longer training times.
4. Interpretability challenges due to the deep and complex architecture, making it difficult to understand internal representations and explain predictions.

11.4. Mobile Net (2017):

The Mobile net network is a lightweight deep neural network proposed by Google for mobile phones and embedded scenarios. Its main feature is to use depth wise separable convolution instead of ordinary convolution, thereby reducing the amount of calculation and improving Computational efficiency of the network.

The classification accuracy of the network on the Image Net dataset has reached 70.8%. In the case of a small loss of accuracy, the amount of calculation is greatly reduced, making it possible for the neural network model to run smoothly on ordinary single-chip computers. [29]

Mobile Net architecture:

Mobile Net is a streamlined architecture that uses depthwise separable convolutions to construct lightweight deep convolutional neural networks and provides an efficient model for mobile and embedded vision applications.

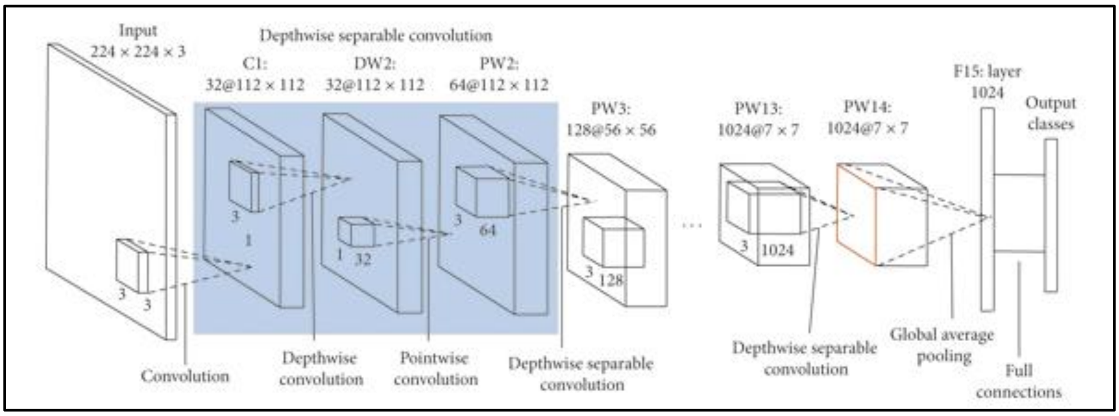

Fig 19. Mobile Net architecture.

The structure of MobileNet is based on depthwise separable filters, as shown in Fig.19.

Advantages of Mobile Nets:

- Efficiency: Mobile Nets are designed to be highly efficient and optimized for mobile devices with limited computational resources.
- Small model size: They have a lightweight architecture, resulting in a small memory footprint, making them suitable for resource-constrained devices.
- Real-time performance: Mobile Nets achieve real-time performance, enabling quick and responsive visual processing.
- Customizability: Developers can adjust the model size and computational requirements to balance accuracy and inference speed based on specific application needs.

Disadvantage of Mobile Nets:

- Reduced accuracy: Compared to larger and more complex CNNs, Mobile Nets may have slightly lower accuracy due to their simplified architecture.
- Limited representation power: They may struggle to capture complex visual patterns and fine-grained details, impacting performance in tasks requiring intricate feature detection.
- Task-specific limitations: Mobile Nets may not be suitable for specialized tasks demanding higher accuracy or relying on specific architectural components not present in Mobile Nets.

12. Conclusion:

In conclusion, computer vision is a challenging field that requires significant effort to delve deep into its intricacies. However, with deep learning methods, we can achieve state-of-the-art results on difficult computer vision problems such as image classification, object detection, and facial recognition.

Among the specialized deep learning algorithms for computer vision, Convolutional Neural Networks (CNNs) have emerged as a dominant approach, thanks to their various architectures introduced in the latter part of this chapter.

Chapter III: Implementation, results and discussion

Chapter III: Implementation, results and discussion

1. Introduction:

In this chapter We will explain the subject on which we worked, the different hardware and software resources that we used, the various experiments that we carried out, we will compare the performance of multiple machine learning algorithms in the context of our project. We will assess factors such as accuracy, precision, recall. etc.

and, finally, a discussion of the results of the evaluation acquired in this chapter.

2. Problematic:

In this work we used CNN architectures to classify Hand sign language, We will use five CNN architectures:

- Custom CNN
- MobileNet
- ResNet50
- EfficientNet (b0)
- Vgg16

3. Configuration used in the implementation:

3.1. Computer :

- ❖ Processor: Intel(R) Xeon(R) CPU @ 2.30GHzwith 2 vCPUs.
- ❖ RAM: 13GB of RAM.
- ❖ Graphics Card: NVIDIA Tesla T4 GPU with 16GB of VRAM.
- ❖ Hard disk: 100 GB.
- ❖ Operating system: Ubuntu 20.04.6 LTS.

3.2. Programming language :

Python was chosen as the programming language for this project due to several key reasons. Firstly, Python is widely recognized as one of the most popular programming languages in the field of data science and machine learning. Its extensive community support and active development ecosystem make it a reliable choice for implementing complex algorithms and models.

Python is a high-level, interpreted programming language known for its simplicity and readability. It was created by Guido van Rossum and first released in 1991. Python emphasizes code readability and uses whitespace indentation to delimit code blocks instead of relying on braces or keywords. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

Python's simplicity and versatility have made it popular among beginners and experienced programmers alike. Its ease of use and readability make it an excellent language for rapid prototyping, scripting, and automating tasks. Python is platform-independent and runs on various operating systems, including Windows, macOS, and Linux.

Overall, Python's strengths lie in its clean syntax, extensive libraries, community support, and broad range of applications, making it a powerful and widely-used programming language in various domains.

Python boasts a rich collection of libraries and tools that greatly facilitated the implementation of the image classification model. The popular deep learning librarys Pytorch, for example, provided a robust framework for constructing, training, and evaluating deep neural networks. Its extensive documentation and active community support enabled the implementation of complex architectures with ease.

An important advantage of using Python for deep learning is the availability of cloud-based platforms such as Google Colab. These platforms provide free access to high-performance GPUs, which are crucial for training deep learning models on large-scale image datasets. The ability to leverage the power of cloud computing significantly speeds up the training process, enabling faster iterations and experimentation.

Overall, the combination of Python's versatility, extensive libraries and tools, and access to cloud platforms like Google Colab made it an ideal choice for training a deep learning model for image classification.

4. Presentation of the datasets:

In this project, we utilized three datasets, which include:

4.1. Sign Language Digits Dataset:

We obtained a dataset of numbers from Kaggle, a popular online platform for data science competitions and datasets. This dataset consists of numerical data that was downloaded and used for analysis and modeling in our project.

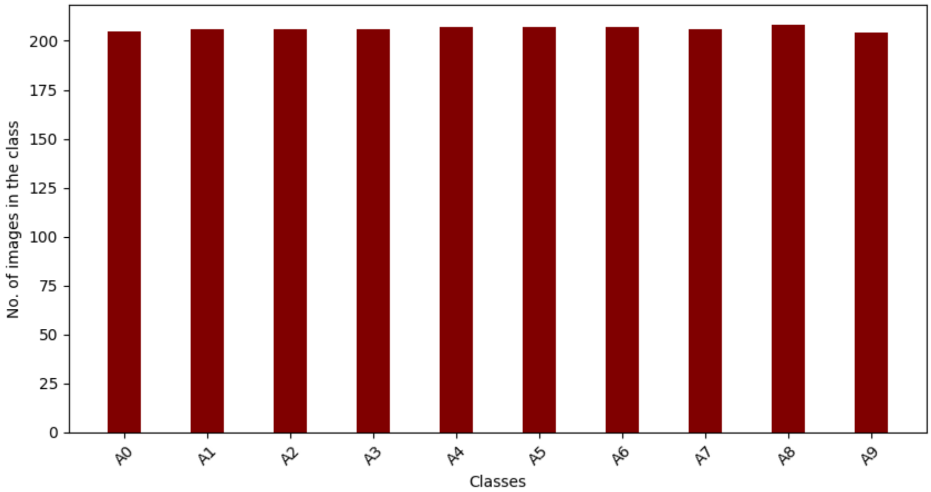

Fig 20. Classes and Number of images in each class (Digits dataset).

Examples of images from Dataset:

Number of classes: 10 (Digits: 0-9), Each class has between 204 and 208 samples. The total data set contains 2062 samples.

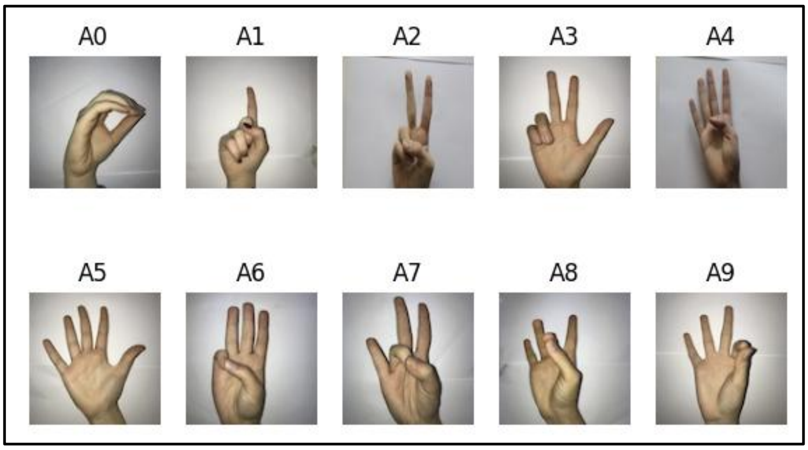

Fig 21. Exampls from Sign Language digits.

4.2. Sign Language Arabic Dataset:

Additionally, we collected a dataset of Arabic signs specifically for this project. This dataset contains images of Arabic signs, which were gathered to train and evaluate our models.

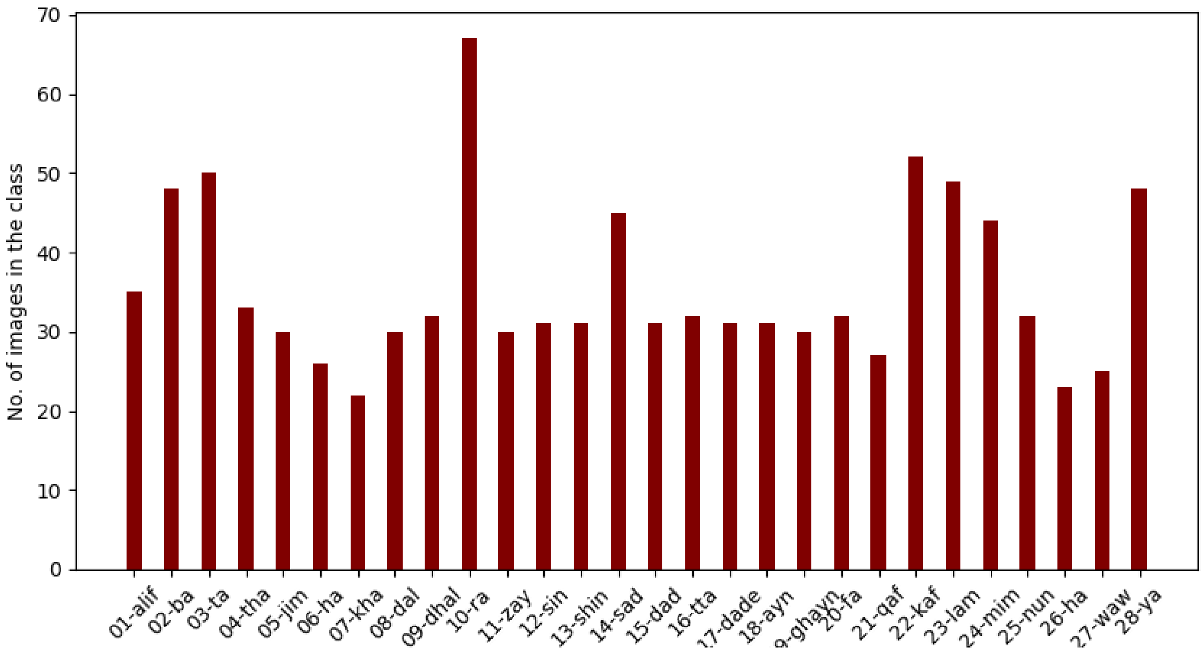

Fig 22. Classes and Number of images in each class (Arabic dataset).

Examples of images from Dataset:

Fig 23. Exampls from Collected Arabic dataset samples.

4.3. Sign Language French Dataset:

Similarly, we also collected a dataset of French signs specifically for this project. This dataset includes images of French signs, which were acquired to train and assess our models.

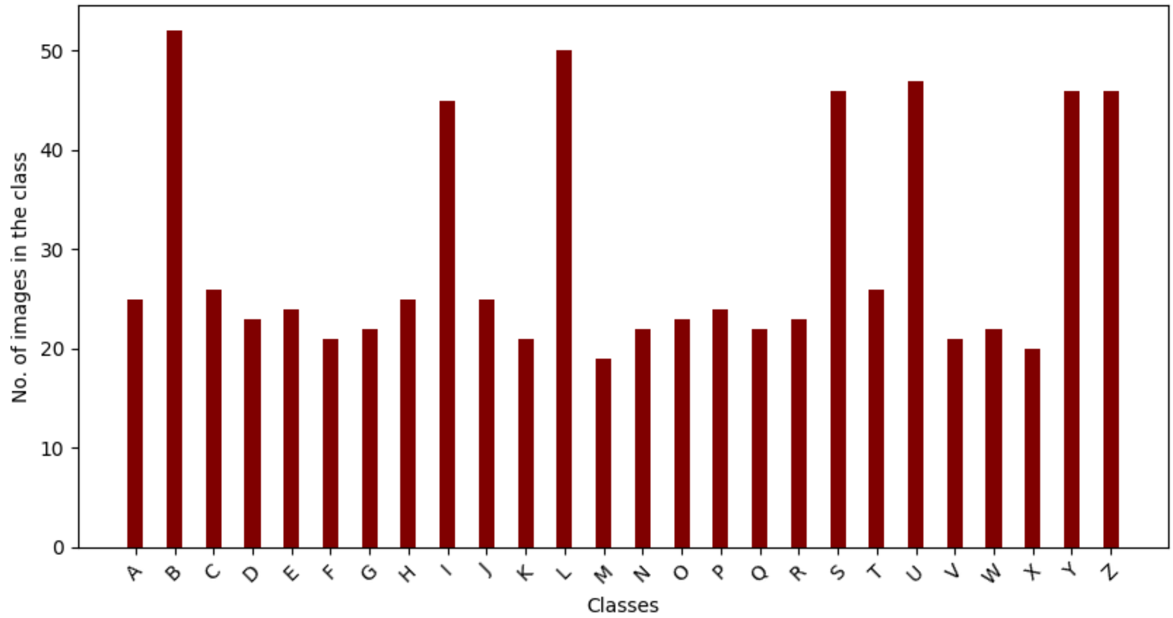

Fig 24. Classes and Number of images in each class (French dataset).

Examples of images from Dataset:

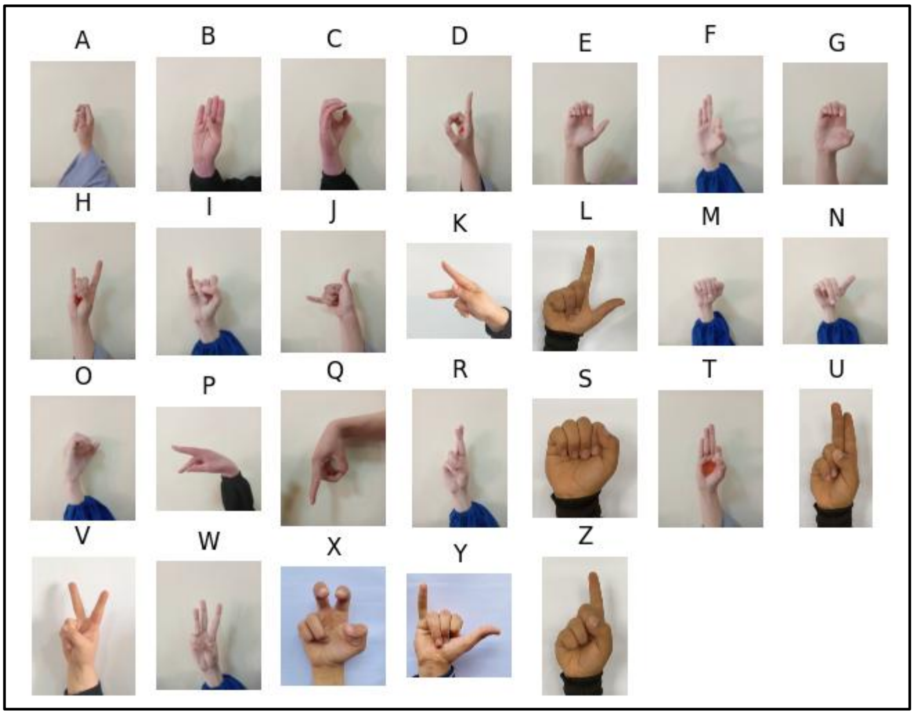

Fig 25. Exampls from Collected French dataset samples.

5. Data pre-processing:

Before starting the training of the chosen architectures, we must go through two data processing stages:

- Crop the bounding box for the hand (region of interest).
- Split of the datasets into two sets [test_set/train_set] 80% for training and 20% for testing.

6. Presentation of the CNN models that we used:

During our experiments we used five models are:

- Model 1: Custom CNN.
- Model 2: ResNet50 pre-trained neural network
- Model 3: Mobilenet V2 pre-trained neural network
- Model 4: VGG16 pre-trained neural network
- Model 5: EfficientNet pre-trained neural network

6.1. Optimizers:

For the compilation of our models we use three optimizers: ADAM, SGD and RMSprop.

6.1.1. ADAM (Adaptive moment estimation):

Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data.

6.1.2. SGD (Stochastic gradient descent):

Deep learning optimizer literature starts with Gradient Descent and the Stochastic Gradient Descent (SGD) is one very widely used version of it. The gradients are not calculated for the loss functions over all data points but over a randomly selected sub-sample. This is why it is also called mini-batch gradient descent sometimes.

6.1.3. RMSprop (Root Mean Square Propagation):

RMSprop is another optimization technique where there is a different learning rate for each parameter. The learning rate is varied by calculating the exponential moving average of the gradient squared and using it to further update the parameter.

6.2. The confusion matrix:

The confusion matrix is a fundamental tool used in the field of machine learning to evaluate the performance of a classification model. It provides a concise summary of the predictions made by the model and how well they align with the actual labels of the data.

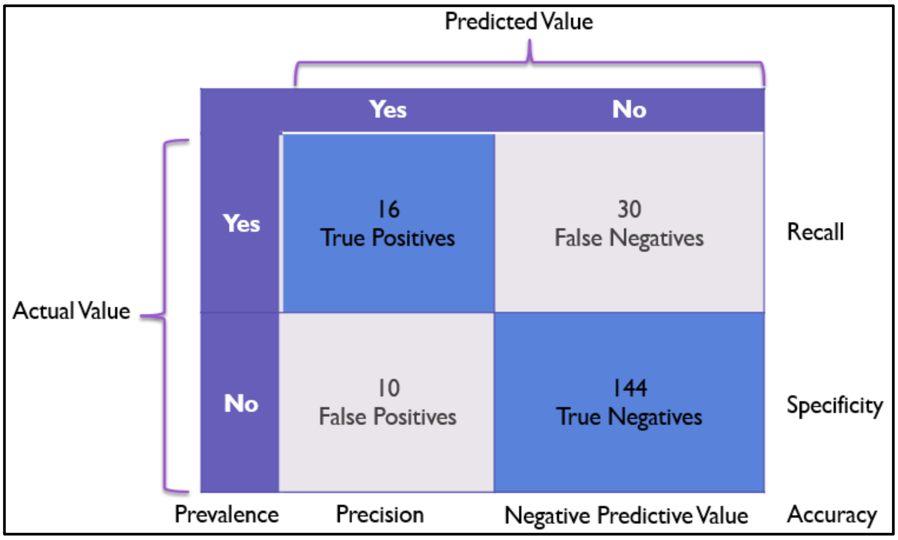

Fig 26. Confusion Matrixe (source:[30])

The confusion matrix is typically a square matrix that displays the counts or proportions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. These values are derived from comparing the predicted labels with the actual labels of the data.

Here's a breakdown of the components of a confusion matrix:

- ❖ True Positive (TP): This represents the number of positive instances that were correctly predicted as positive by the model.
- ❖ True Negative (TN): This indicates the number of negative instances that were correctly predicted as negative by the model.
- ❖ False Positive (FP): This refers to the number of negative instances that were incorrectly predicted as positive by the model. Also known as a Type I error.
- ❖ False Negative (FN): This represents the number of positive instances that were incorrectly predicted as negative by the model. Also known as a Type II error.

The confusion matrix allows us to calculate various performance metrics that provide insights into the model's effectiveness, such as accuracy, precision, recall, and F1 score. These metrics are calculated using combinations of the values in the confusion matrix.

Accuracy: measures the overall correctness of the model and is calculated as (TP + TN) / (TP + TN + FP + FN). It gives the proportion of correct predictions out of the total number of instances.

Precision: (also known as positive predictive value) measures the model's ability to correctly identify positive instances and is calculated as TP / (TP + FP). It indicates the proportion of correctly predicted positive instances out of all instances predicted as positive.

Recall: (also known as sensitivity or true positive rate) measures the model's ability to correctly identify positive instances from all actual positive instances and is calculated as TP / (TP + FN). It represents the proportion of correctly predicted positive instances out of all actual positive instances.

F1 score: is the harmonic mean of precision and recall and provides a balanced evaluation metric. It is calculated as 2 * (precision * recall) / (precision + recall).

By analyzing the confusion matrix and the associated metrics, we can gain a better understanding of a model's strengths and weaknesses, particularly in terms of its ability to correctly classify different classes and avoid misclassifications. This information can guide us in refining and improving the model's performance.

7. The training stages of our models:

7.1. The first step:

In this stage, we trained all models for 10 Epoches with learning rate 0.001 and change each time the Batch size (8, 16, 32) and the optimizers (ADAM, SGD, RMSprop).

Table 1. Number of images in each dataset.

Datasets	**Total images**	**Train set**	**Validation set**
Digits	2062	1649	413
French	766	612	154
Arabic	997	797	200

7.1.1. The result from digits dataset:

Table 2. Detailed results from digits dataset.

Dataset	model	Batch size	Optimizer	Train Acc	Valid Acc	Selected
Digits	Custom CNN	8	ADAM	87.93	91.04	X
		8	SGD	33.96	17.19	
		8	RMSprop	57.55	68.77	
		16	ADAM	68.89	81.11	
		16	SGD	21.89	10.17	
		16	RMSprop	16.13	18.16	
		32	ADAM	83.99	70.22	
		32	SGD	18.98	26.63	
		32	RMSprop	30.14	28.09	
	ResNet	8	ADAM	99.76	99.76	X
		8	SGD	95.69	93.22	
		8	RMSprop	98.06	99.52	
		16	ADAM	99.70	96.13	
		16	SGD	92.30	87.17	
		16	RMSprop	98.36	11.86	
		32	ADAM	99.33	98.55	
		32	SGD	80.41	60.53	
		32	RMSprop	96.06	93.95	
	MobileNet V2	8	ADAM	99.39	99.76	X
		8	SGD	82.11	89.35	
		8	RMSprop	99.88	82.57	
		16	ADAM	98.97	98.79	
		16	SGD	69.50	68.04	
		16	RMSprop	98.85	63.68	
		32	ADAM	100.00	100.00	
		32	SGD	49.48	55.21	
		32	RMSprop	99.64	100.00	
	Vgg 16	8	ADAM	10.43	08.47	
		8	SGD	09.76	10.65	
		8	RMSprop	10.43	08.47	
		16	ADAM	09.64	08.47	
		16	SGD	09.76	10.65	X
		16	RMSprop	10.31	08.47	
		32	ADAM	10.37	08.47	
		32	SGD	10.13	09.20	
		32	RMSprop	10.19	09.20	
	EfficientNet (b0)	8	ADAM	98.67	99.76	
		8	SGD	90.60	96.37	
		8	RMSprop	98.48	85.71	
		16	ADAM	98.79	99.52	
		16	SGD	85.32	86.92	
		16	RMSprop	98.36	99.76	
		32	ADAM	99.94	99.76	X
		32	SGD	72.59	73.12	
		32	RMSprop	98.97	84.26	

7.1.2. The result from French dataset:

Table 3. Detailed results from French dataset.

Dataset	model	Batch size	Optimizer	Train Acc	Valid Acc	selected
French	Custom CNN	8	ADAM	35.46	35.06	X
			SGD	10.62	05.19	
			RMSprop	08.50	05.19	
		16	ADAM	62.91	33.12	
			SGD	07.52	08.44	
			RMSprop	08.99	06.49	
		32	ADAM	08.99	07.14	
			SGD	09.31	05.19	
			RMSprop	24.35	13.64	
	ResNet	8	ADAM	98.37	90.26	X
			SGD	50.82	34.42	
			RMSprop	94.61	88.31	
		16	ADAM	97.55	90.26	
			SGD	25.33	11.69	
			RMSprop	85.46	64.94	
		32	ADAM	97.55	80.52	
			SGD	06.86	01.30	
			RMSprop	90.85	46.75	
	MobileNet V2	8	ADAM	93.95	89.61	
			SGD	19.44	19.48	
			RMSprop	93.46	92.21	X
		16	ADAM	97.06	91.56	
			SGD	14.54	14.29	
			RMSprop	93.79	64.29	
		32	ADAM	97.71	90.26	
			SGD	11.76	13.64	
			RMSprop	99.02	88.31	
	Vgg 16	8	ADAM	06.86	03.25	
			SGD	03.92	00.65	
			RMSprop	06.86	06.49	X
		16	ADAM	05.72	06.49	
			SGD	03.92	00.65	
			RMSprop	06.54	06.49	
		32	ADAM	06.54	03.90	
			SGD	03.92	00.65	
			RMSprop	06.86	03.25	
	EfficientNet (b0)	8	ADAM	96.41	92.21	X
			SGD	44.93	35.71	
			RMSprop	94.93	82.47	
		16	ADAM	94.93	87.01	
			SGD	14.71	11.69	
			RMSprop	97.71	87.01	
		32	ADAM	98.04	86.36	
			SGD	02.29	03.90	
			RMSprop	95.92	79.87	

7.1.3. The result from Arabic dataset:

Table 4. Detailed results from Arabic dataset.

Dataset	model	Batch size	Optimizer	Train Acc	Valid Acc	selected
Arabic	Custom CNN	8	ADAM	75.03	67.50	X
			SGD	09.28	13.50	
			RMSprop	39.40	34.00	
		16	ADAM	44.04	43.00	
			SGD	09.03	10.00	
			RMSprop	07.53	05.50	
		32	ADAM	39.27	33.00	
			SGD	08.66	10.00	
			RMSprop	41.78	37.50	
	ResNet	8	ADAM	99.25	93.50	X
			SGD	67.50	57.00	
			RMSprop	95.73	89.00	
		16	ADAM	99.25	91.50	
			SGD	38.02	25.50	
			RMSprop	93.73	82.50	
		32	ADAM	99.25	88.50	
			SGD	15.31	12.00	
			RMSprop	94.23	65.50	
	MobileNet V2	8	ADAM	96.61	87.00	
			SGD	19.95	18.00	
			RMSprop	97.62	93.00	X
		16	ADAM	98.62	92.50	
			SGD	16.06	18.00	
			RMSprop	99.12	90.00	
		32	ADAM	98.87	91.50	
			SGD	13.05	10.50	
			RMSprop	97.49	86.00	
	Vgg 16	8	ADAM	06.65	07.00	
			SGD	03.76	02.50	
			RMSprop	06.65	07.00	
		16	ADAM	06.65	07.00	
			SGD	03.76	02.50	
			RMSprop	06.65	07.00	
		32	ADAM	05.14	05.50	
			SGD	92.60	79.00	X
			RMSprop	04.89	04.50	
	EfficientNet (b0)	8	ADAM	99.50	97.00	
			SGD	53.58	50.50	
			RMSprop	95.48	93.50	
		16	ADAM	98.12	87.00	
			SGD	20.83	14.00	
			RMSprop	97.99	95.00	
		32	ADAM	99.37	95.50	
			SGD	02.76	04.00	
			RMSprop	99.75	97.00	X

7.2. The second step:

In this stage, we will train models that clearly provided satisfactory results in the first stage (Custom CNN, MobileNet, ResNet, EfficientNet). for 20 Epoches using small learning rate (0.0001) to avoid overfitting, and the best **Batch size** and best **Optimizer** from the first-stage as follow:

Table 5. Hyperparameters for second step

Dataset	Model	Batch size	Optimizer
Digits	Custom CNN	8	ADAM
	ResNet	8	ADAM
	MobileNet	8	ADAM
	EfficientNet	32	ADAM
French	Custom CNN	8	ADAM
	ResNet	8	ADAM
	MobileNet	8	RMSprop
	EfficientNet	8	ADAM
Arabic	Custom CNN	8	ADAM
	ResNet	8	ADAM
	MobileNet	8	RMSprop
	EfficientNet	32	RMSprop

7.2.1. The result from Digits dataset:

After training the models using the specified parameters, we analyzed the training and validation loss to evaluate their performance. Here are the results:

Custom CNN:

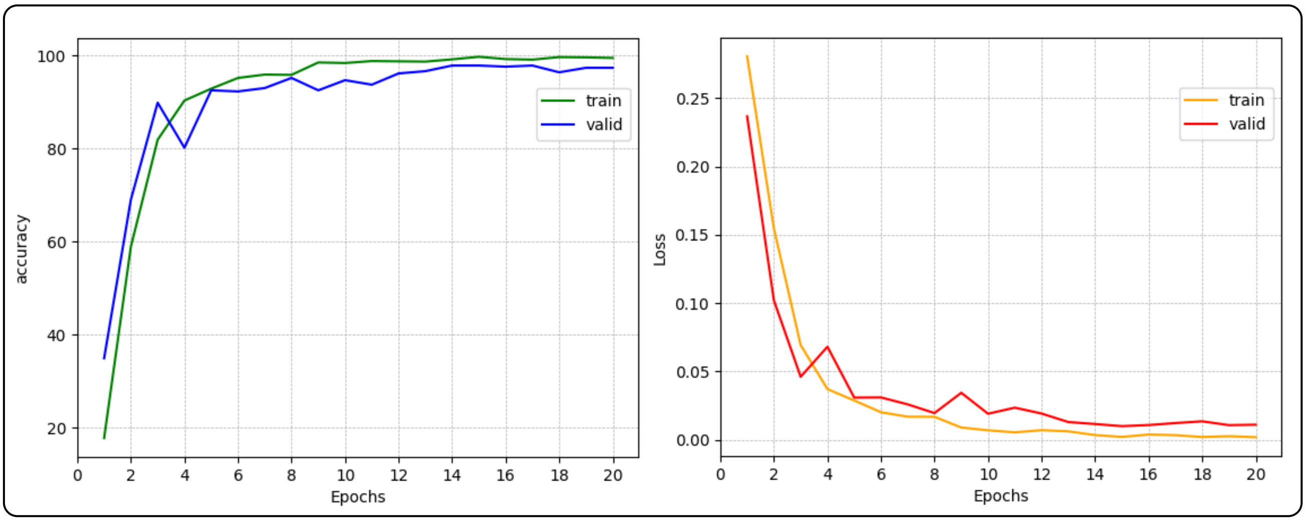

Fig 27. Accuracy and Loss achieved: Custom CNN with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A0	0.98	0.96	0.97	45
A1	0.98	0.98	0.98	47
A2	0.96	1.00	0.98	46
A3	1.00	0.97	0.98	31
A4	0.95	1.00	0.97	39
A5	0.94	1.00	0.97	34
A6	1.00	0.96	0.98	49
A7	0.98	0.95	0.96	43
A8	0.95	0.97	0.96	37
A9	1.00	0.95	0.98	42
accuracy			0.97	413

Fig 28. Classification report: Custom CNN with batch size: 8 and ADAM optimizer.

Confusion Matrix:

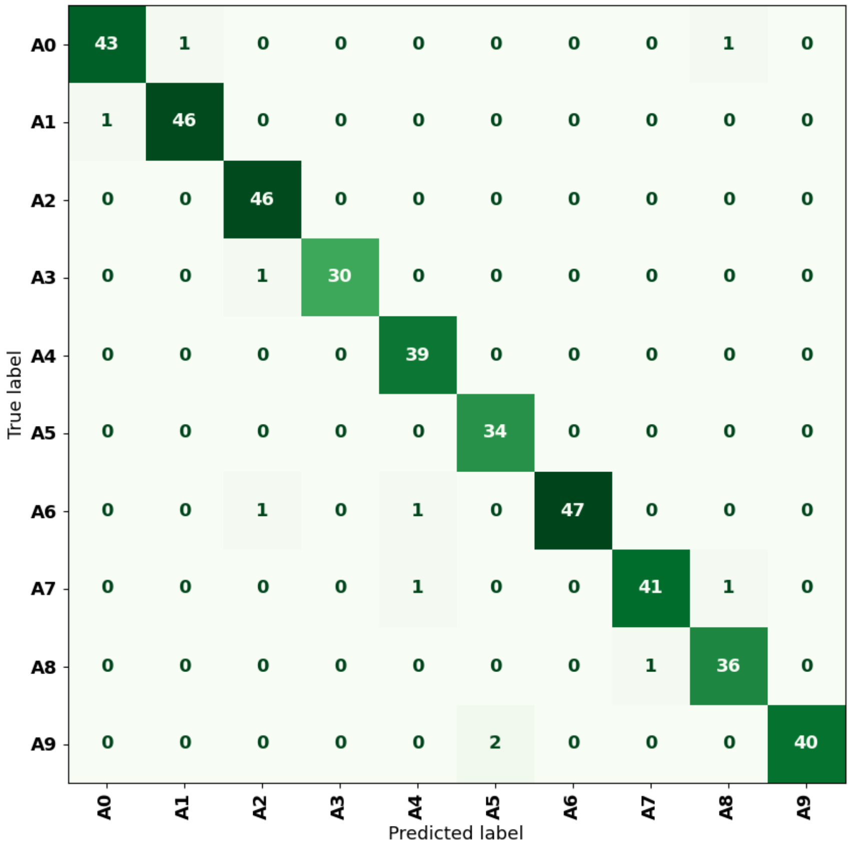

Fig 29. Confusion matrix: Custom CNN with batch size: 8 and ADAM optimizer.

- ResNet:

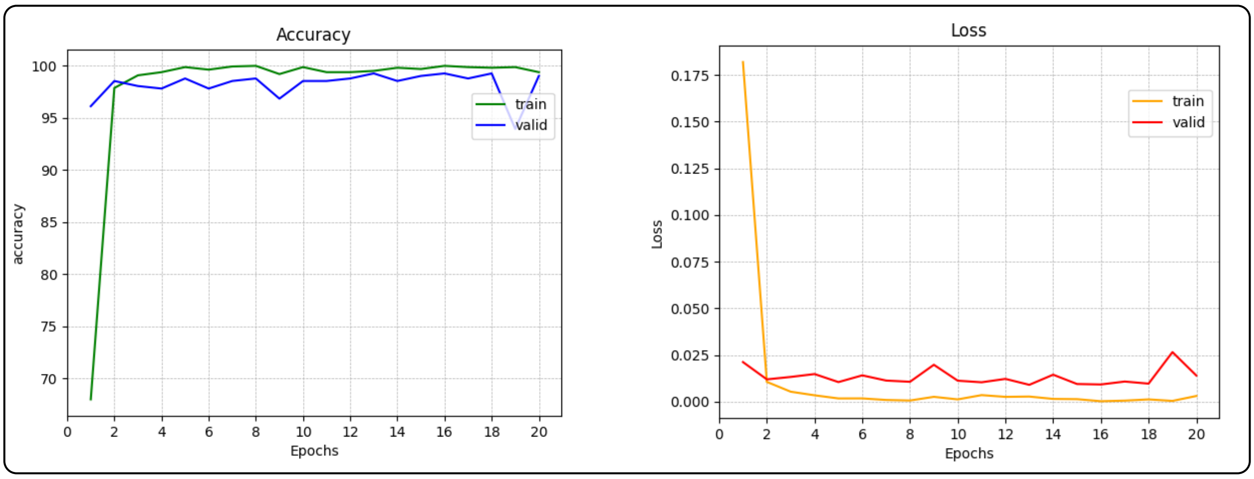

Fig 30. Accuracy and Loss achieved: ResNet with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A0	1.00	1.00	1.00	39
A1	1.00	1.00	1.00	49
A2	1.00	0.97	0.98	32
A3	0.97	0.95	0.96	39
A4	1.00	1.00	1.00	43
A5	0.97	1.00	0.98	30
A6	1.00	1.00	1.00	50
A7	0.98	0.98	0.98	42
A8	0.98	1.00	0.99	41
A9	1.00	1.00	1.00	48
accuracy			0.99	413

Fig 31. Classification report: ResNet with batch size: 8 and ADAM optimizer.

Confusion Matrix:

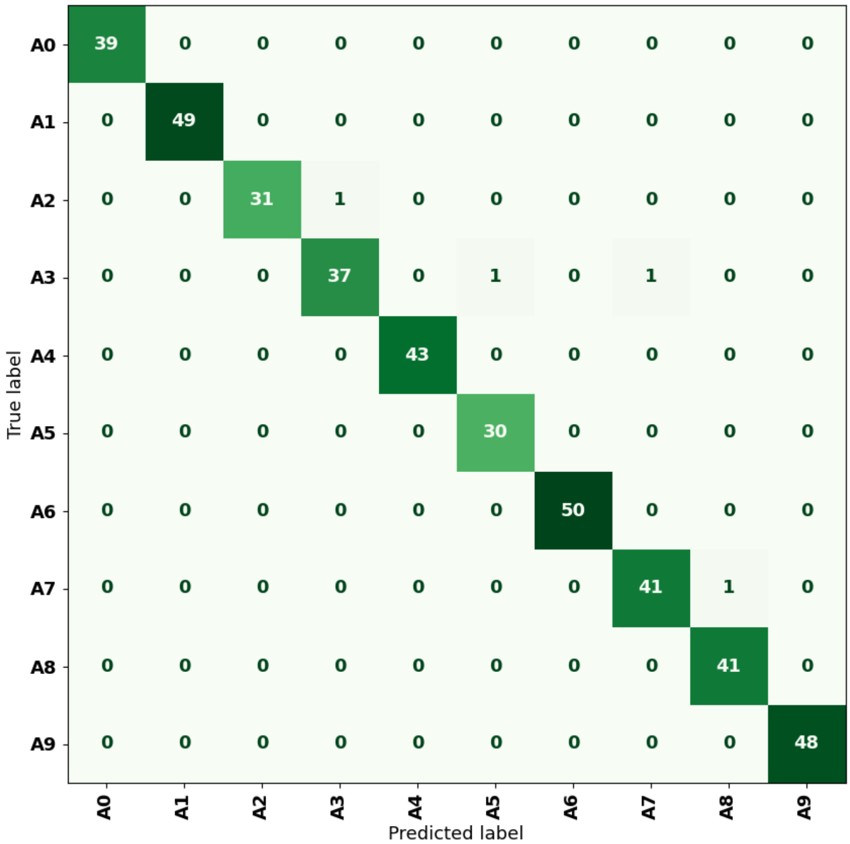

Fig 32. Confusion Matrix: ResNet with batch size: 8 and ADAM optimizer.

- **MobileNet** with batch size: **8** and **ADAM** optimizer.

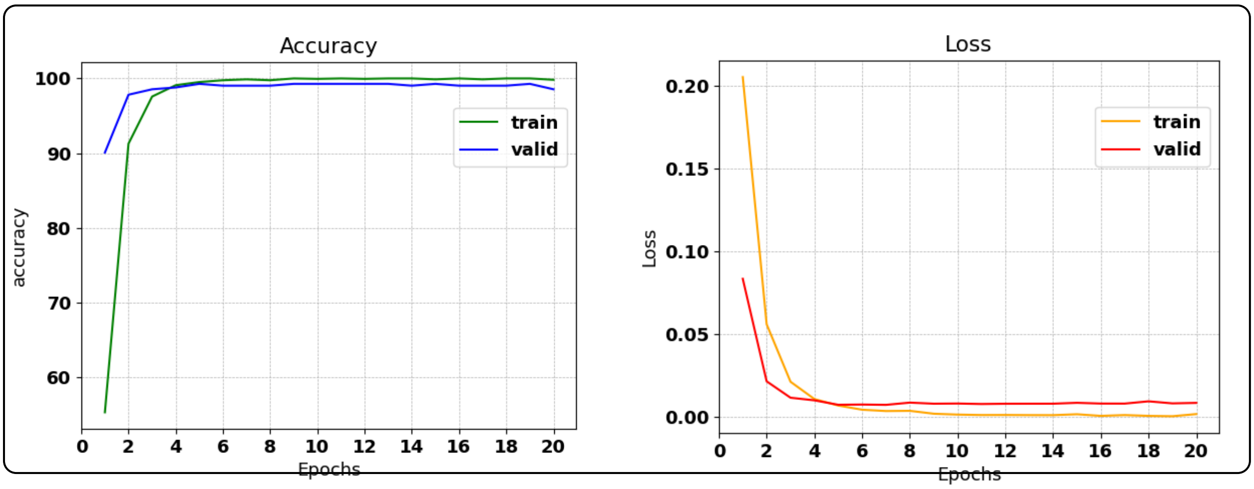

Fig 33. Accuracy and Loss achieved: MobileNet with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A0	1.00	1.00	1.00	39
A1	1.00	1.00	1.00	49
A2	0.94	0.97	0.95	32
A3	0.97	0.92	0.95	39
A4	0.98	1.00	0.99	43
A5	1.00	1.00	1.00	30
A6	1.00	1.00	1.00	50
A7	0.95	1.00	0.98	42
A8	1.00	0.95	0.97	41
A9	1.00	1.00	1.00	48
accuracy			0.99	413

Fig 34. Classification report: MobileNet with batch size: 8 and ADAM optimizer.

Confusion Matrix:

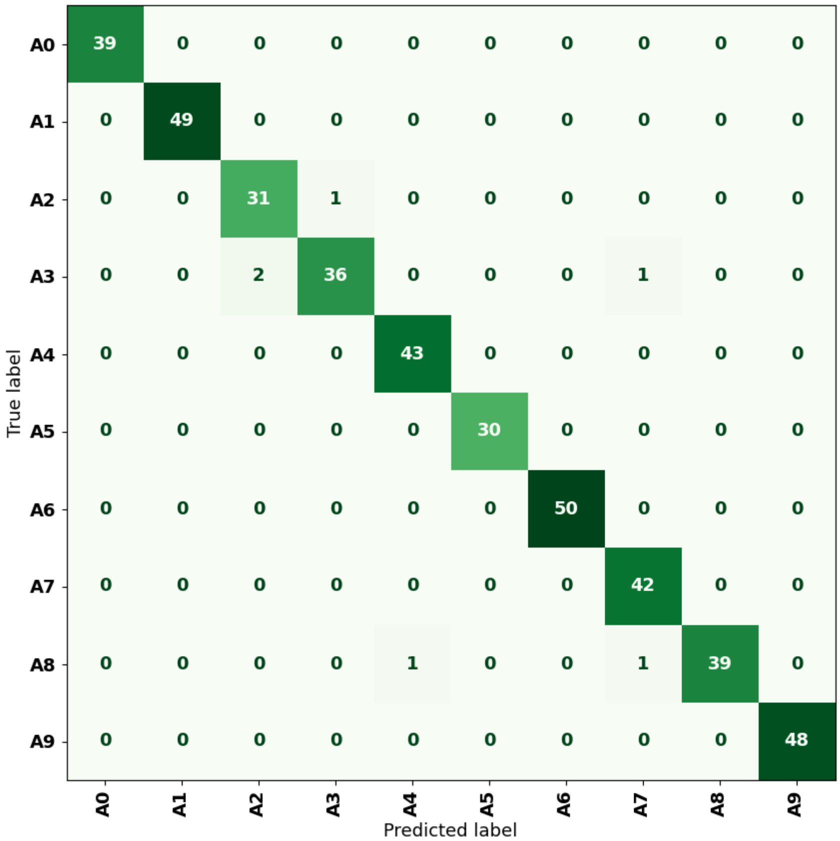

Fig 35. Confusion Matrix: MobileNet with batch size: 8 and ADAM optimizer.

- EfficientNet:

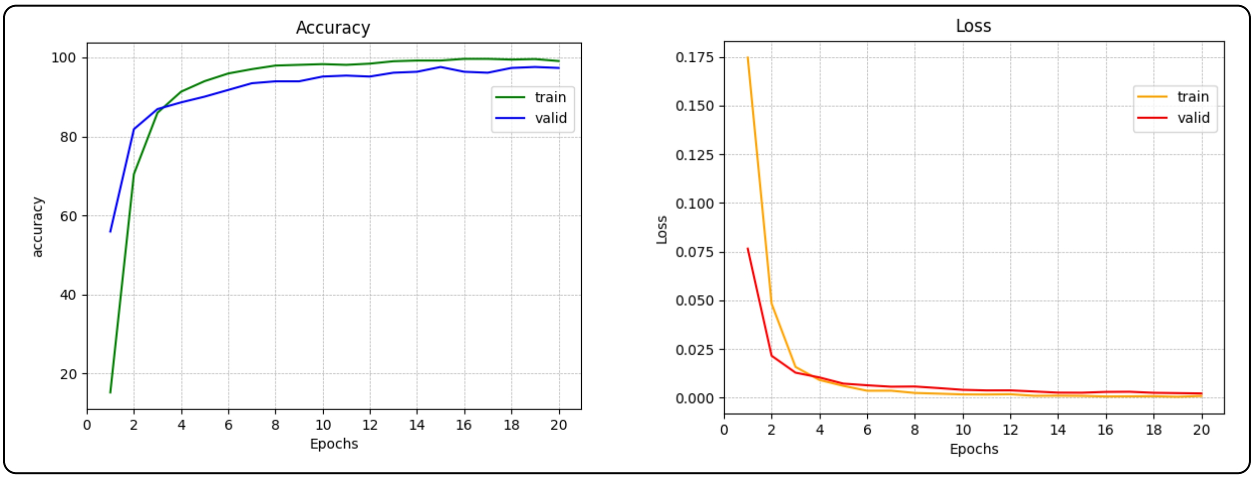

Fig 36. Accuracy and Loss achieved: EfficientNet with batch size: 32 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A0	1.00	1.00	1.00	47
A1	1.00	0.98	0.99	49
A2	0.95	0.95	0.95	42
A3	1.00	1.00	1.00	31
A4	0.98	0.91	0.94	45
A5	1.00	1.00	1.00	34
A6	0.89	0.95	0.92	43
A7	0.97	0.97	0.97	39
A8	0.95	0.98	0.96	42
A9	1.00	1.00	1.00	41
accuracy			0.97	413

Fig 37. Classification report: EfficientNet with batch size: 32 and ADAM optimizer.

Confusion Matrix:

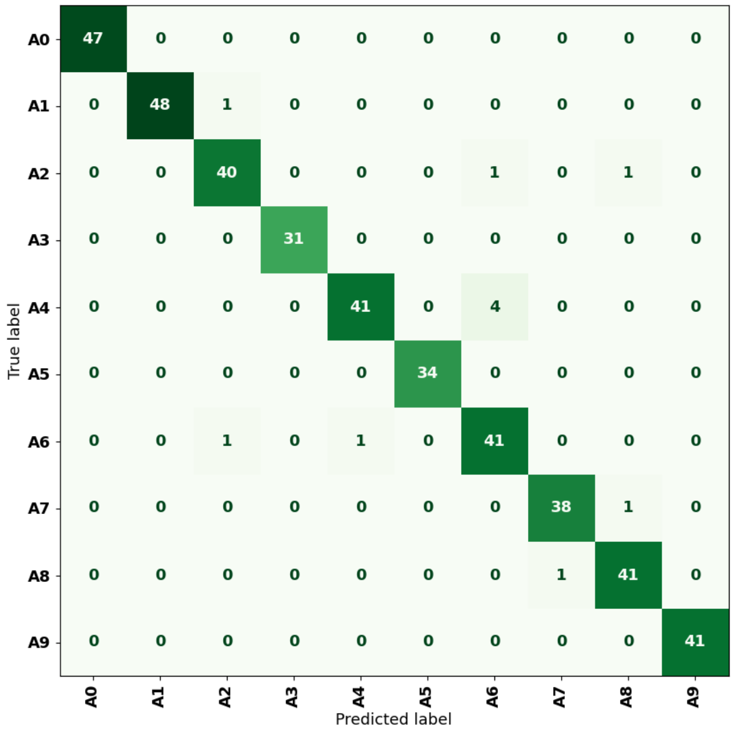

Fig 38. Confusion Matrix: EfficientNet with batch size: 32 and ADAM optimizer.

From the comparison, we can observe the following:

- All four models demonstrate high accuracy, with ResNet and MobileNet achieving the highest accuracy of 0.99.
- ResNet consistently shows high precision, recall, and F1-scores across all classes, indicating excellent performance overall.
- MobileNet also performs well, with high precision, recall, and F1-scores, although some classes have slightly lower scores compared to ResNet .
- Model 1 (Custom CNN) and Model 4 (EfficientNet) exhibit slightly lower performance in terms of precision, recall, and F1-scores compared to the other two models.
- Model 4 (EfficientNet) shows the lowest scores among the four models, particularly in precision and recall for certain classes.

7.2.2. The result from French dataset:

After training the models using the specified parameters, we analyzed the training and validation loss to evaluate their performance. Here are the results:

- **Custom CNN**:

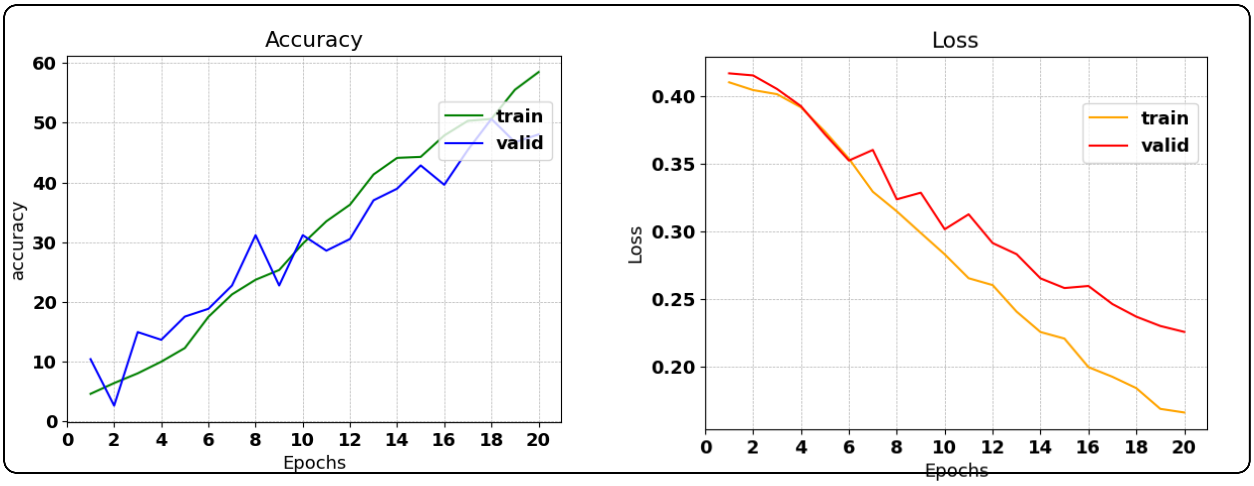

Fig 39. Accuracy and Loss achieved: Custom CNN with batch size: 8 and ADAM optimizer.

Confusion Matrix:

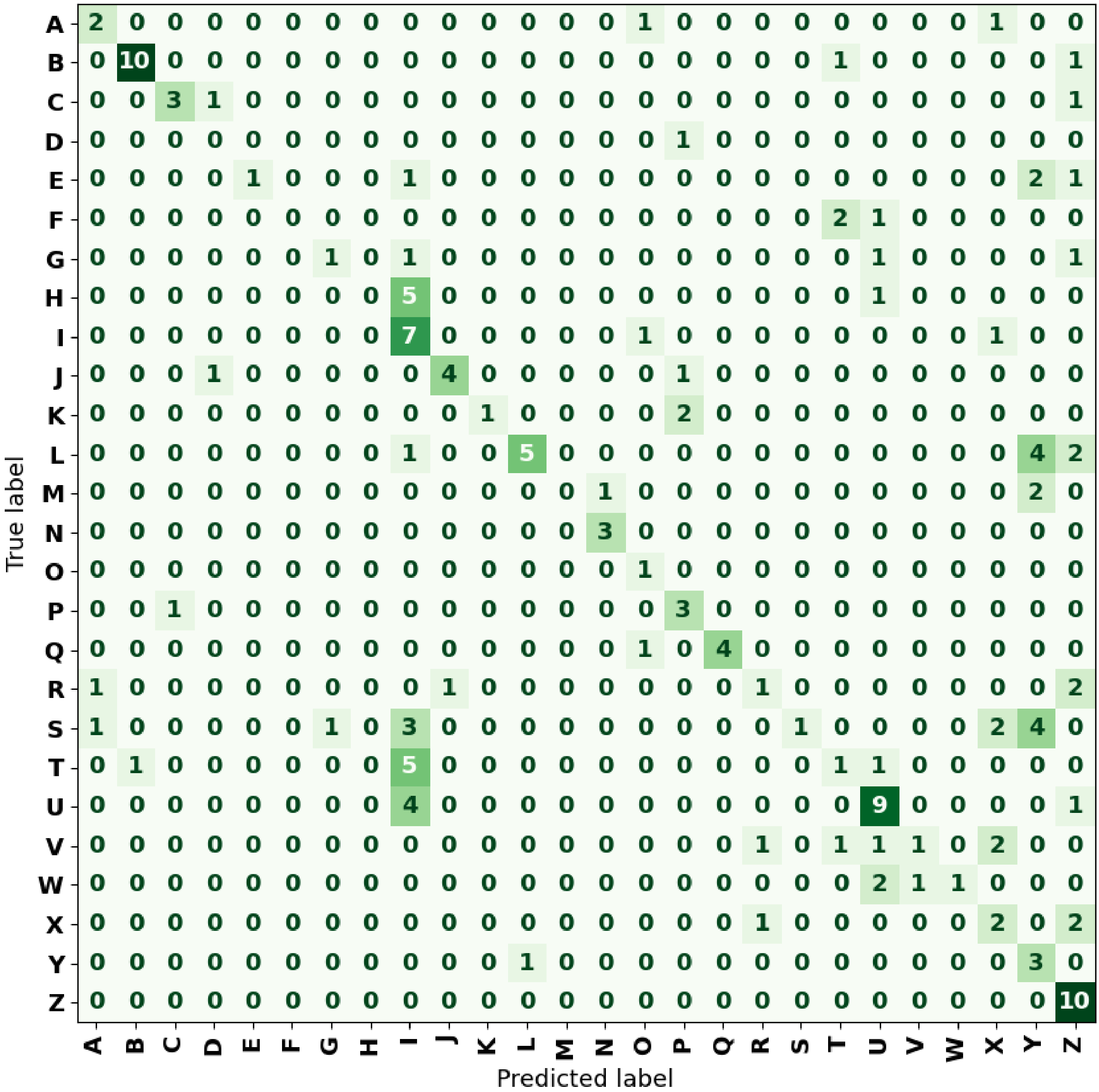

Fig 40. Confusion Matrix: Custom CNN with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A	0.50	0.50	0.50	4
B	0.91	0.83	0.87	12
C	0.75	0.60	0.67	5
D	0.00	0.00	0.00	1
E	1.00	0.20	0.33	5
F	0.00	0.00	0.00	3
G	0.50	0.25	0.33	4
H	0.00	0.00	0.00	6
I	0.26	0.78	0.39	9
J	0.80	0.67	0.73	6
K	1.00	0.33	0.50	3
L	0.83	0.42	0.56	12
M	0.00	0.00	0.00	3
N	0.75	1.00	0.86	3
O	0.25	1.00	0.40	1
P	0.43	0.75	0.55	4
Q	1.00	0.80	0.89	5
R	0.33	0.20	0.25	5
S	1.00	0.08	0.15	12
T	0.20	0.12	0.15	8
U	0.56	0.64	0.60	14
V	0.50	0.17	0.25	6
W	1.00	0.25	0.40	4
X	0.25	0.40	0.31	5
Y	0.20	0.75	0.32	4
Z	0.48	1.00	0.65	10
accuracy			0.48	154

Fig 41. Classification report: Custom CNN with batch size: 8 and ADAM optimizer.

- ResNet:

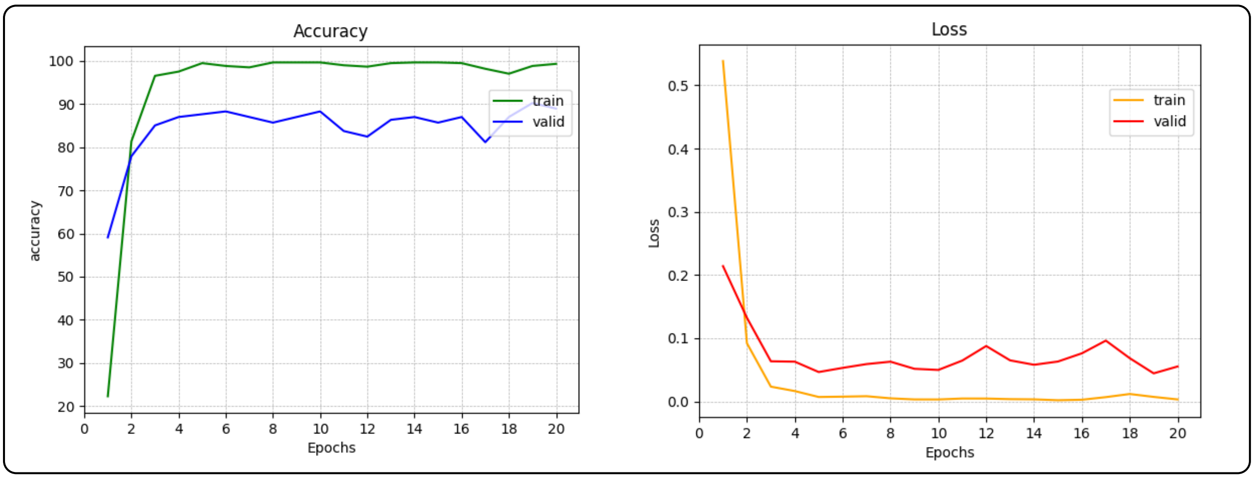

Fig 42. Accuracy and Loss achieved: ResNet with batch size: 8 and ADAM optimizer.

Confusion Matrix:

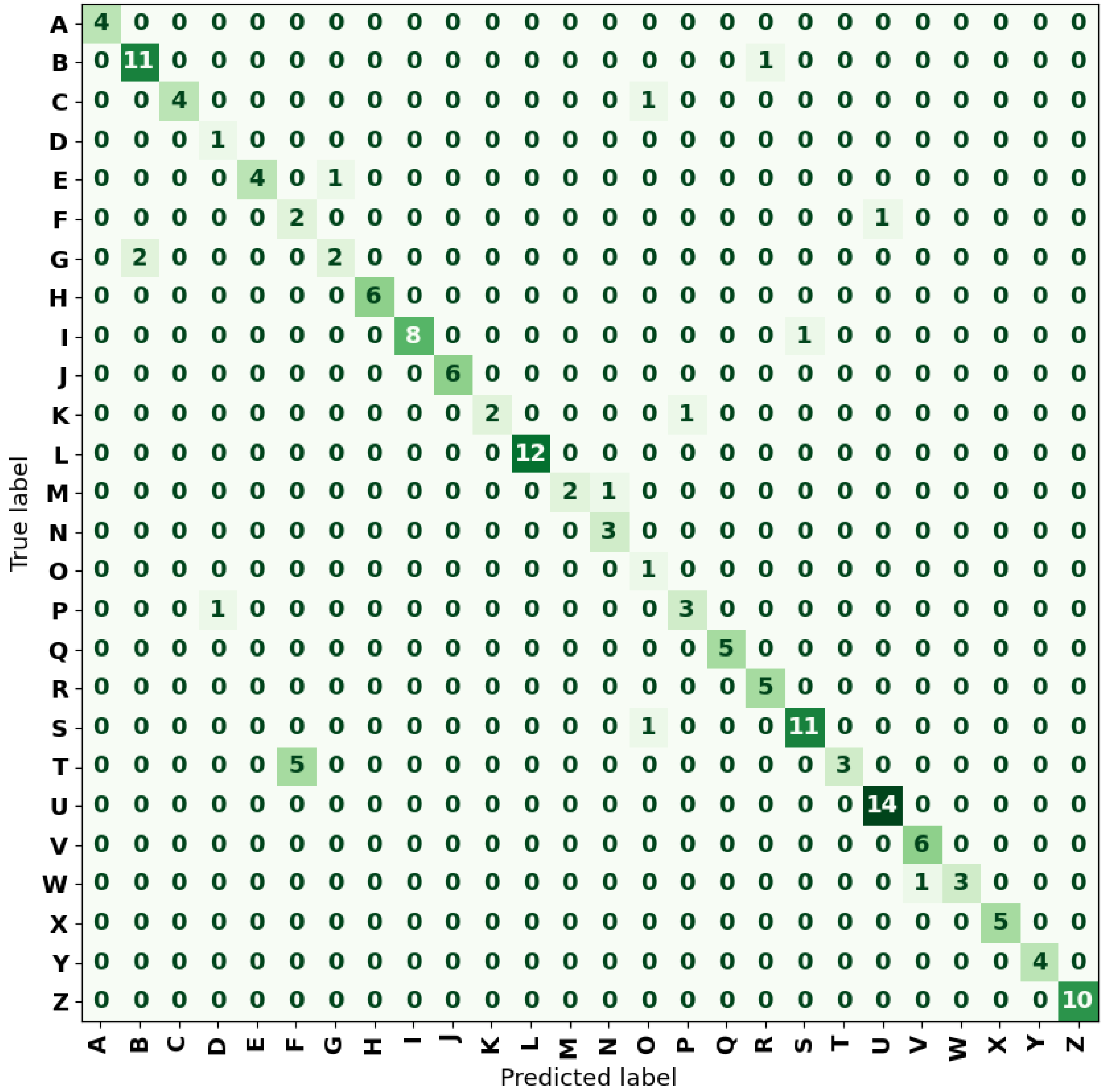

Fig 43. Confusion Matrix: ResNet with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A	1.00	1.00	1.00	4
B	0.85	0.92	0.88	12
C	1.00	0.80	0.89	5
D	0.50	1.00	0.67	1
E	1.00	0.80	0.89	5
F	0.29	0.67	0.40	3
G	0.67	0.50	0.57	4
H	1.00	1.00	1.00	6
I	1.00	0.89	0.94	9
J	1.00	1.00	1.00	6
K	1.00	0.67	0.80	3
L	1.00	1.00	1.00	12
M	1.00	0.67	0.80	3
N	0.75	1.00	0.86	3
O	0.33	1.00	0.50	1
P	0.75	0.75	0.75	4
Q	1.00	1.00	1.00	5
R	0.83	1.00	0.91	5
S	0.92	0.92	0.92	12
T	1.00	0.38	0.55	8
U	0.93	1.00	0.97	14
V	0.86	1.00	0.92	6
W	1.00	0.75	0.86	4
X	1.00	1.00	1.00	5
Y	1.00	1.00	1.00	4
Z	1.00	1.00	1.00	10
accuracy			0.89	154

Fig 44. Classification report: ResNet with batch size: 8 and ADAM optimizer.

- **MobileNet** with batch size: **8** and **RMSprop** optimizer.

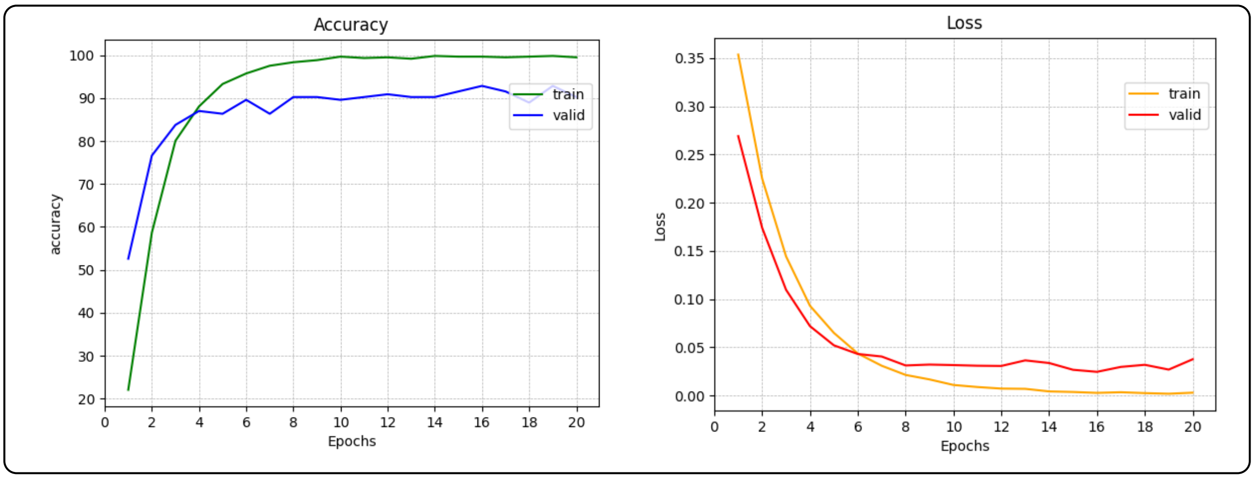

Fig 45. Accuracy and Loss achieved: MobileNet with batch size: 8 and RMSprop optimizer.

Confusion Matrix:

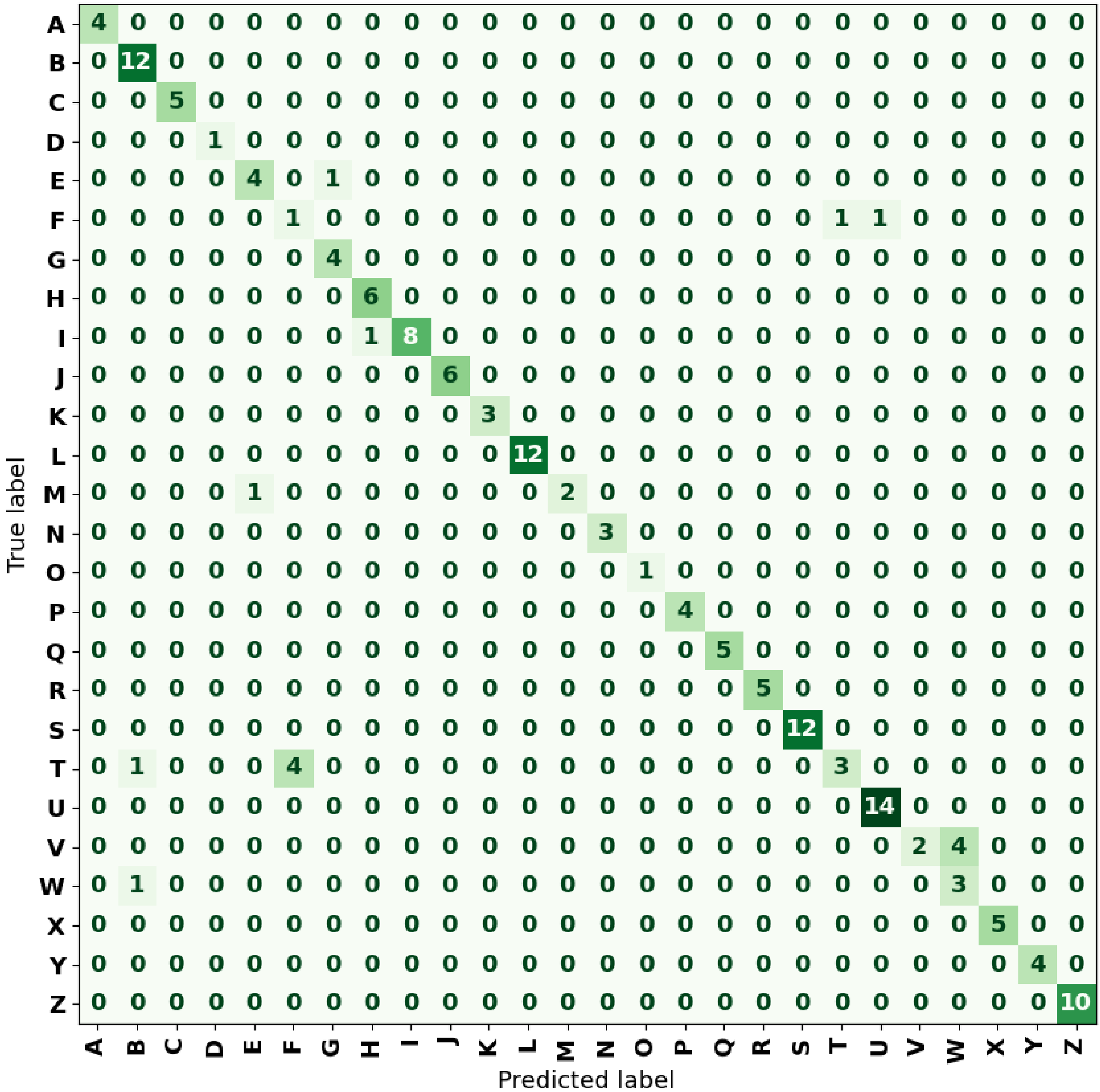

Fig 46. Confusion Matrix: MobileNet with batch size: 8 and RMSprop optimizer.

Classification report:

	precision	recall	f1-score	support
A	1.00	1.00	1.00	4
B	0.86	1.00	0.92	12
C	1.00	1.00	1.00	5
D	1.00	1.00	1.00	1
E	0.80	0.80	0.80	5
F	0.20	0.33	0.25	3
G	0.80	1.00	0.89	4
H	0.86	1.00	0.92	6
I	1.00	0.89	0.94	9
J	1.00	1.00	1.00	6
K	1.00	1.00	1.00	3
L	1.00	1.00	1.00	12
M	1.00	0.67	0.80	3
N	1.00	1.00	1.00	3
O	1.00	1.00	1.00	1
P	1.00	1.00	1.00	4
Q	1.00	1.00	1.00	5
R	1.00	1.00	1.00	5
S	1.00	1.00	1.00	12
T	0.75	0.38	0.50	8
U	0.93	1.00	0.97	14
V	1.00	0.33	0.50	6
W	0.43	0.75	0.55	4
X	1.00	1.00	1.00	5
Y	1.00	1.00	1.00	4
Z	1.00	1.00	1.00	10
accuracy			0.90	154

Fig 47. Classification report: MobileNet with batch size: 8 and RMSprop optimizer.

- EfficientNet:

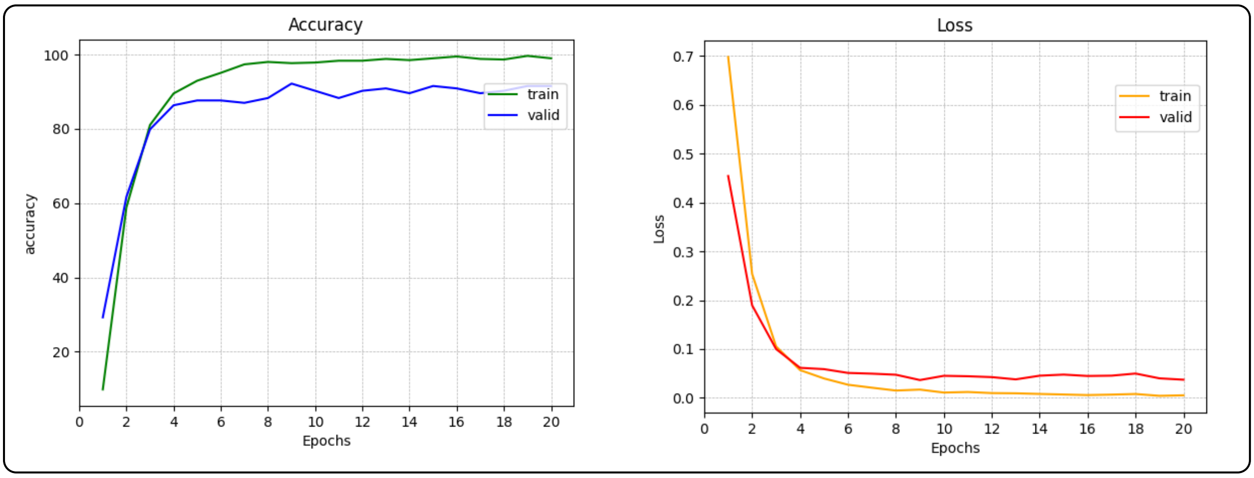

Fig 48. Accuracy and Loss achieved: EfficientNet with batch size: 32 and ADAM optimizer.

Confusion Matrix:

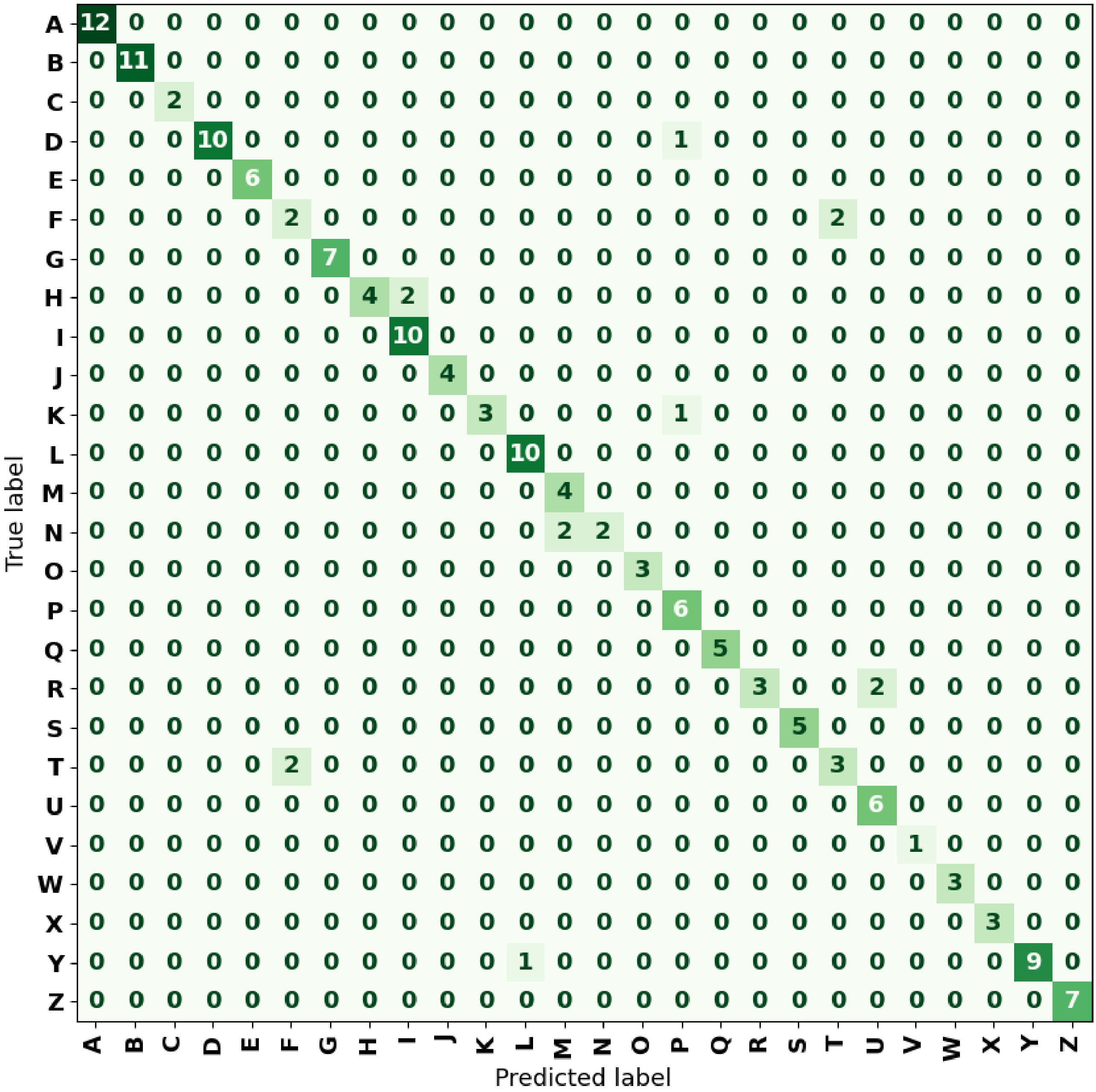

Fig 49. Confusion Matrix: EfficientNet with batch size: 32 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
A	1.00	1.00	1.00	12
B	1.00	1.00	1.00	11
C	1.00	1.00	1.00	2
D	1.00	0.91	0.95	11
E	1.00	1.00	1.00	6
F	0.50	0.50	0.50	4
G	1.00	1.00	1.00	7
H	1.00	0.67	0.80	6
I	0.83	1.00	0.91	10
J	1.00	1.00	1.00	4
K	1.00	0.75	0.86	4
L	0.91	1.00	0.95	10
M	0.67	1.00	0.80	4
N	1.00	0.50	0.67	4
O	1.00	1.00	1.00	3
P	0.75	1.00	0.86	6
Q	1.00	1.00	1.00	5
R	1.00	0.60	0.75	5
S	1.00	1.00	1.00	5
T	0.60	0.60	0.60	5
U	0.75	1.00	0.86	6
V	1.00	1.00	1.00	1
W	1.00	1.00	1.00	3
X	1.00	1.00	1.00	3
Y	1.00	0.90	0.95	10
Z	1.00	1.00	1.00	7
accuracy			0.92	154

Fig 50. Classification report: EfficientNet with batch size: 32 and ADAM optimizer.

From the comparison, we can observe the following:

- Model 4 (EfficientNet) achieves the highest accuracy (0.92) among all four models, followed closely by Model 3 (MobileNet) with an accuracy of 0.90.
- Model 1 (Custom CNN) has the lowest accuracy (0.48) among the four models.
- Model 4 (EfficientNet) also shows the highest average precision (0.92) and average recall (0.91), indicating strong performance in correctly identifying positive and negative instances.
- Model 2 (ResNet) demonstrates competitive performance across all metrics, with an accuracy of 0.89 and an average F1-score of 0.87.
- Model 1 (Custom CNN) struggles with low precision, recall, and F1-scores across multiple classes, resulting in its low accuracy.

7.2.3. The result from Arabic dataset:

After training the models using the specified parameters, we analyzed the training and validation loss to evaluate their performance. Here are the results:

- Custom CNN:

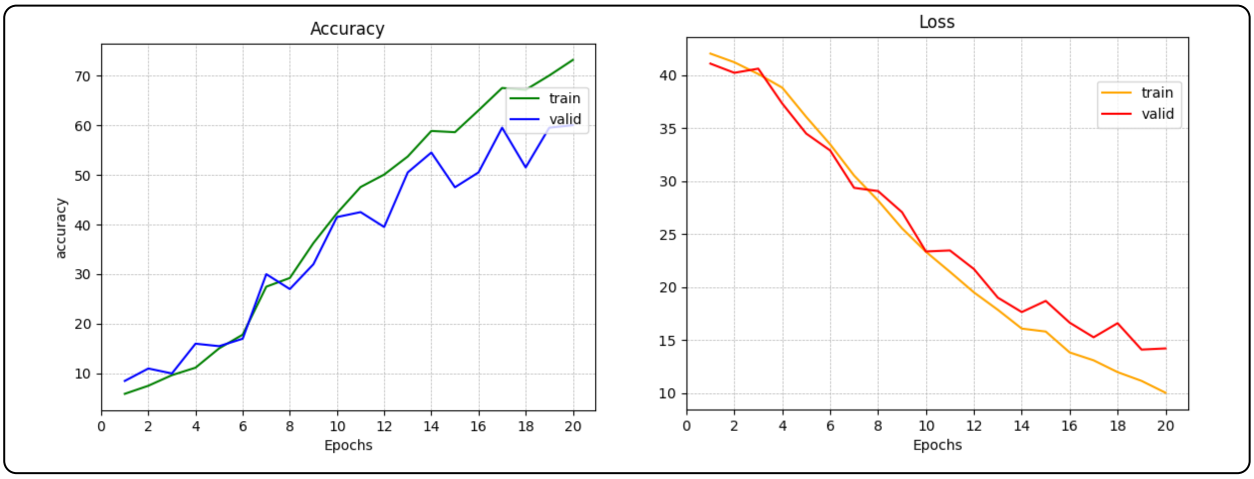

Fig 51. Accuracy and Loss achieved: Custom CNN with batch size: 8 and ADAM optimizer.

Confusion Matrix:

	01-alif	02-ba	03-ta	04-tha	05-jim	06-ha	07-kha	08-dal	09-dhal	10-ra	11-zay	12-sin	13-shin	14-sad	15-dad	16-tta	17-dade	18-ayn	19-ghayn	20-fa	21-qaf	22-kaf	23-lam	24-mim	25-nun	26-ha	27-waw	28-ya
01-alif	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
02-ba	0	8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
03-ta	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
04-tha	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
05-jim	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
06-ha	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07-kha	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
08-dal	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
09-dhal	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10-ra	0	0	0	0	1	0	0	0	0	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
11-zay	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0
12-sin	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
13-shin	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
14-sad	0	0	1	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	1	0	0	0	1	0	0	0	0
15-dad	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0
16-tta	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
17-dade	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	1	0	0	0	0	0
18-ayn	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0
19-ghayn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
20-fa	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	3	0	0	0	0	0	0	0	0
21-qaf	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	2	2	1	0	1	0	0	0	0
22-kaf	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0
23-lam	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0	0	0
24-mim	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	6	0	0	0	1
25-nun	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0
26-ha	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0
27-waw	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
28-ya	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	3

Fig 52. Confusion Matrix: Custom CNN with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
01-alif	0.88	1.00	0.93	7
02-ba	0.73	0.80	0.76	10
03-ta	0.71	0.91	0.80	11
04-tha	0.00	0.00	0.00	11
05-jim	0.75	0.60	0.67	5
06-ha	0.50	0.20	0.29	5
07-kha	0.17	0.50	0.25	2
08-dal	0.67	0.67	0.67	3
09-dhal	1.00	0.80	0.89	5
10-ra	0.88	0.82	0.85	17
11-zay	0.60	0.50	0.55	6
12-sin	0.12	0.20	0.15	5
13-shin	1.00	0.83	0.91	6
14-sad	0.47	0.75	0.58	12
15-dad	1.00	0.20	0.33	5
16-tta	0.83	0.83	0.83	6
17-dade	0.00	0.00	0.00	5
18-ayn	1.00	0.12	0.22	8
19-ghayn	0.42	1.00	0.59	5
20-fa	0.43	0.43	0.43	7
21-qaf	0.67	0.22	0.33	9
22-kaf	0.33	0.73	0.46	11
23-lam	0.70	0.70	0.70	10
24-mim	0.55	0.67	0.60	9
25-nun	1.00	0.80	0.89	5
26-ha	0.60	0.75	0.67	4
27-waw	0.57	1.00	0.73	4
28-ya	0.75	0.43	0.55	7
accuracy			0.60	200

Fig 53. Classification report: Custom CNN with batch size: 8 and ADAM optimizer.

- **ResNet**:

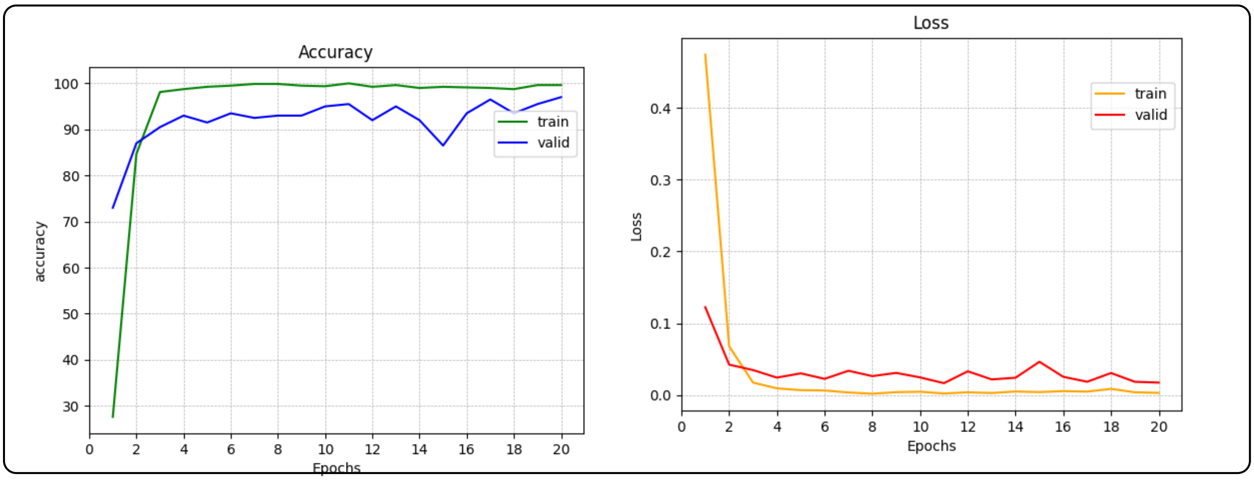

Fig 54. Accuracy and Loss achieved: ResNet with batch size: 8 and ADAM optimizer.

Confusion Matrix:

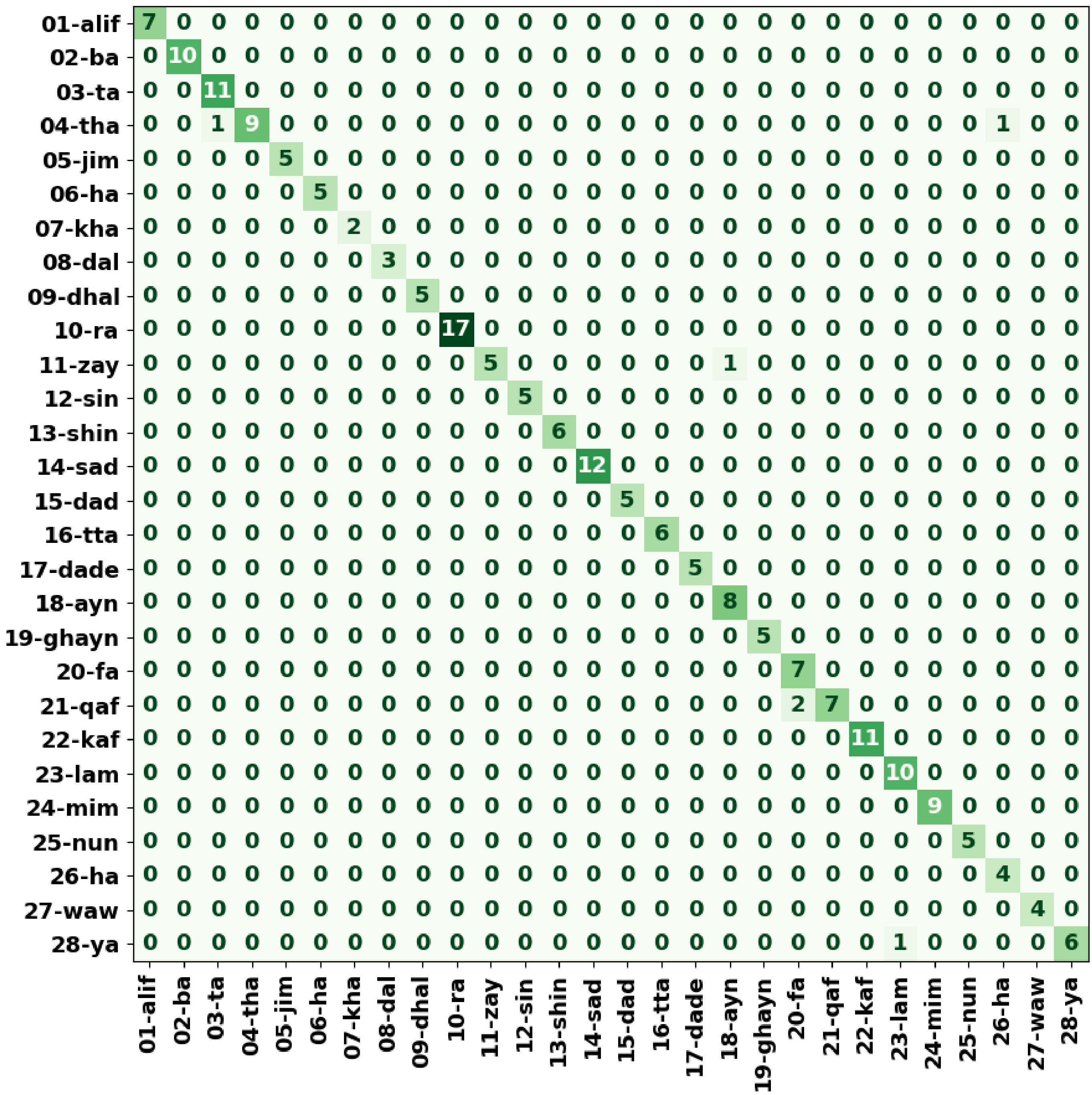

Fig 55. Confusion Matrix: ResNet with batch size: 8 and ADAM optimizer.

Classification report:

	precision	recall	f1-score	support
01-alif	1.00	1.00	1.00	7
02-ba	1.00	1.00	1.00	10
03-ta	0.92	1.00	0.96	11
04-tha	1.00	0.82	0.90	11
05-jim	1.00	1.00	1.00	5
06-ha	1.00	1.00	1.00	5
07-kha	1.00	1.00	1.00	2
08-dal	1.00	1.00	1.00	3
09-dhal	1.00	1.00	1.00	5
10-ra	1.00	1.00	1.00	17
11-zay	1.00	0.83	0.91	6
12-sin	1.00	1.00	1.00	5
13-shin	1.00	1.00	1.00	6
14-sad	1.00	1.00	1.00	12
15-dad	1.00	1.00	1.00	5
16-tta	1.00	1.00	1.00	6
17-dade	1.00	1.00	1.00	5
18-ayn	0.89	1.00	0.94	8
19-ghayn	1.00	1.00	1.00	5
20-fa	0.78	1.00	0.88	7
21-qaf	1.00	0.78	0.88	9
22-kaf	1.00	1.00	1.00	11
23-lam	0.91	1.00	0.95	10
24-mim	1.00	1.00	1.00	9
25-nun	1.00	1.00	1.00	5
26-ha	0.80	1.00	0.89	4
27-waw	1.00	1.00	1.00	4
28-ya	1.00	0.86	0.92	7
accuracy			0.97	200

Fig 56. Classification report: ResNet with batch size: 8 and ADAM optimizer.

- MobileNet:

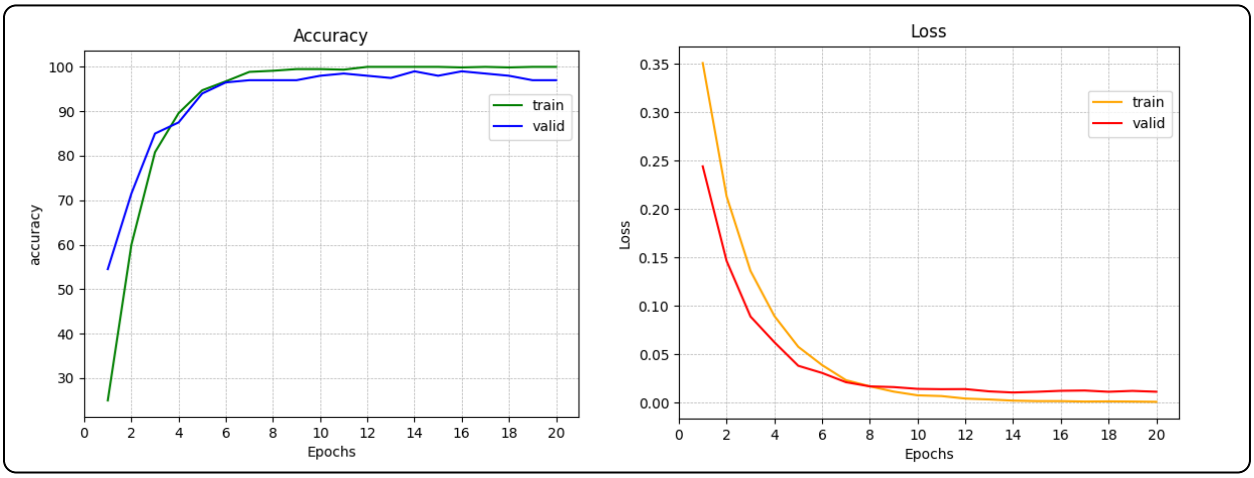

Fig 57. Accuracy and Loss achieved: MobileNet with batch size: 8 and RMSprop optimizer.

Confusion Matrix:

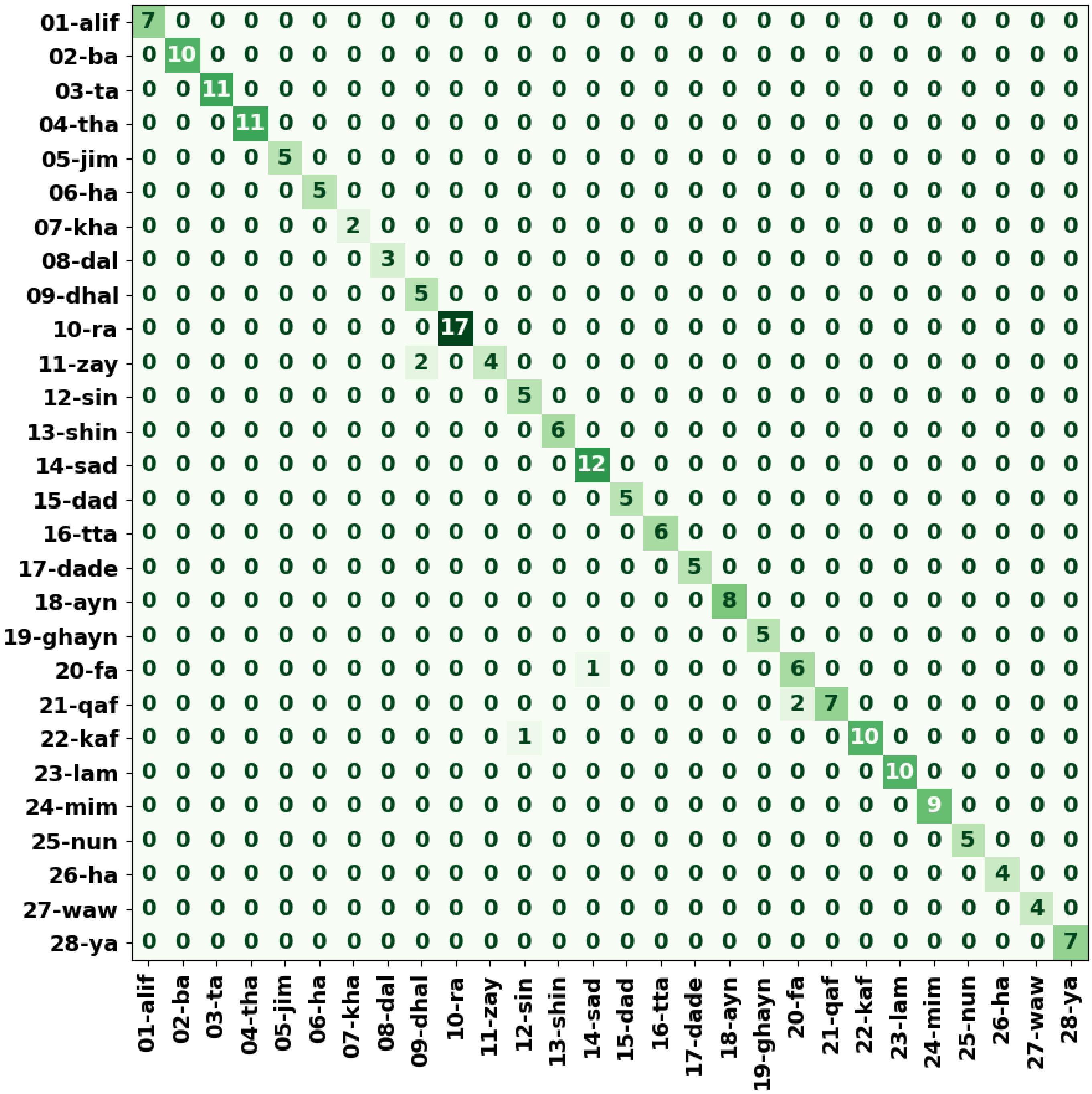

Fig 58. Confusion Matrix: MobileNet with batch size: 8 and RMSprop optimizer.

Classification report:

	precision	recall	f1-score	support
01-alif	1.00	1.00	1.00	7
02-ba	1.00	1.00	1.00	10
03-ta	1.00	1.00	1.00	11
04-tha	1.00	1.00	1.00	11
05-jim	1.00	1.00	1.00	5
06-ha	1.00	1.00	1.00	5
07-kha	1.00	1.00	1.00	2
08-dal	1.00	1.00	1.00	3
09-dhal	0.71	1.00	0.83	5
10-ra	1.00	1.00	1.00	17
11-zay	1.00	0.67	0.80	6
12-sin	0.83	1.00	0.91	5
13-shin	1.00	1.00	1.00	6
14-sad	0.92	1.00	0.96	12
15-dad	1.00	1.00	1.00	5
16-tta	1.00	1.00	1.00	6
17-dade	1.00	1.00	1.00	5
18-ayn	1.00	1.00	1.00	8
19-ghayn	1.00	1.00	1.00	5
20-fa	0.75	0.86	0.80	7
21-qaf	1.00	0.78	0.88	9
22-kaf	1.00	0.91	0.95	11
23-lam	1.00	1.00	1.00	10
24-mim	1.00	1.00	1.00	9
25-nun	1.00	1.00	1.00	5
26-ha	1.00	1.00	1.00	4
27-waw	1.00	1.00	1.00	4
28-ya	1.00	1.00	1.00	7
accuracy			0.97	200

Fig 59. Classification report: MobileNet with batch size: 8 and RMSprop optimizer.

- EfficientNet:

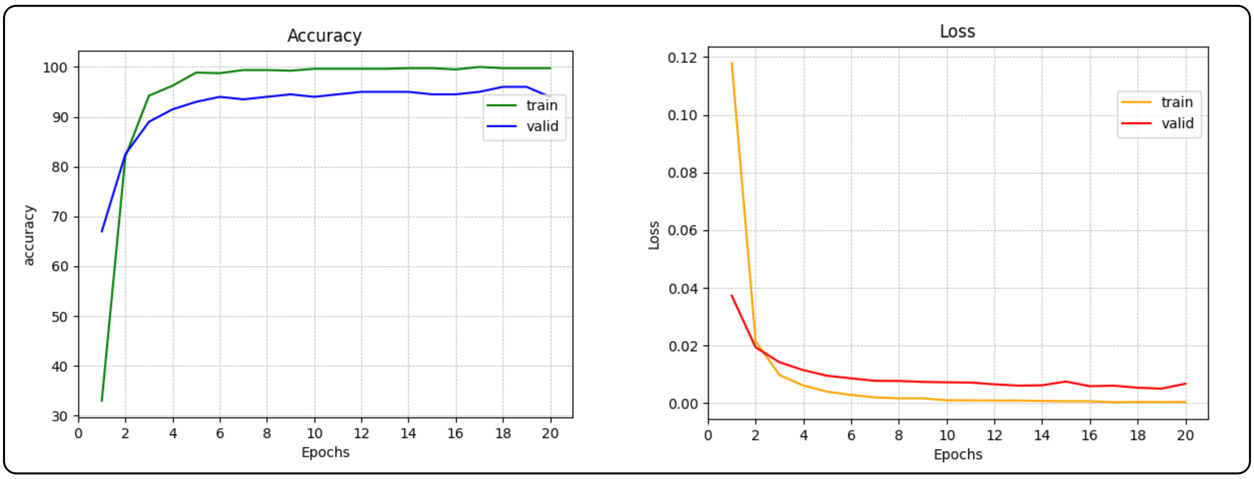

Fig 60. Accuracy and Loss achieved: EfficientNet with batch size: 8 and RMSprop optimizer.

Confusion Matrix:

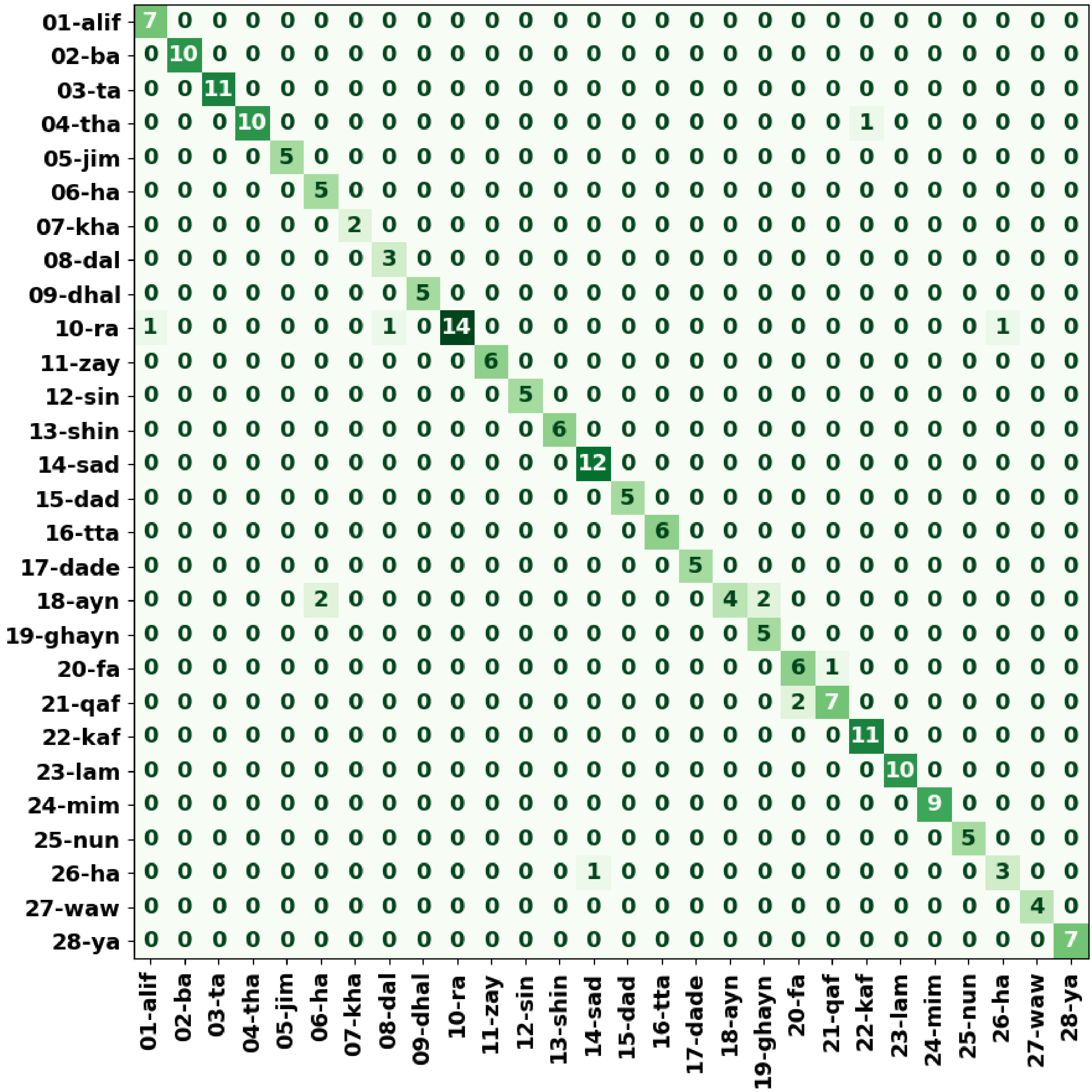

Fig 61. Confusion Matrix: EfficientNet with batch size: 8 and RMSprop optimizer.

Classification report:

	precision	recall	f1-score	support
01-alif	0.88	1.00	0.93	7
02-ba	1.00	1.00	1.00	10
03-ta	1.00	1.00	1.00	11
04-tha	1.00	0.91	0.95	11
05-jim	1.00	1.00	1.00	5
06-ha	0.71	1.00	0.83	5
07-kha	1.00	1.00	1.00	2
08-dal	0.75	1.00	0.86	3
09-dhal	1.00	1.00	1.00	5
10-ra	1.00	0.82	0.90	17
11-zay	1.00	1.00	1.00	6
12-sin	1.00	1.00	1.00	5
13-shin	1.00	1.00	1.00	6
14-sad	0.92	1.00	0.96	12
15-dad	1.00	1.00	1.00	5
16-tta	1.00	1.00	1.00	6
17-dade	1.00	1.00	1.00	5
18-ayn	1.00	0.50	0.67	8
19-ghayn	0.71	1.00	0.83	5
20-fa	0.75	0.86	0.80	7
21-qaf	0.88	0.78	0.82	9
22-kaf	0.92	1.00	0.96	11
23-lam	1.00	1.00	1.00	10
24-mim	1.00	1.00	1.00	9
25-nun	1.00	1.00	1.00	5
26-ha	0.75	0.75	0.75	4
27-waw	1.00	1.00	1.00	4
28-ya	1.00	1.00	1.00	7
accuracy			0.94	200

Fig 62. Classification report: EfficientNet with batch size: 8 and RMSprop optimizer.

From the comparison, we can observe the following:

- Model 2 (ResNet) and Model 3 (MobileNet) achieve the highest accuracy (0.97) among all four models, indicating strong overall performance.
- Model 1 (Custom CNN) exhibits lower accuracy (0.60) compared to the other models.
- Model 2 (ResNet) and Model 3 (MobileNet) demonstrate the highest average precision, recall, and F1-scores, indicating their effectiveness in correctly identifying positive and negative instances across multiple classes.
- Model 4 (EfficientNet) also shows competitive performance with an accuracy of 0.94 and relatively high precision, recall, and F1-scores.

8. Conclusion:

In conclusion, this chapter focused on the implementation, results, and discussion of the project. The configurations used in the implementation were specified, including the computer and programming language utilized. The datasets used in the project were presented, including the Sign Language Digits Dataset, Sign Language Arabic Dataset, and Sign Language French Dataset.

The training stages of the models were explained, with the first step involving training all models for 10 epochs, while also conducting hyperparameter tuning. This included varying batch sizes (8, 16, 32) and experimenting with different optimizers to identify the optimal settings for each model.

The second step involved training selected models that yielded satisfactory results in the first stage, namely Custom CNN, MobileNet, ResNet, and EfficientNet.

In this stage, the models were trained for 20 epochs with a lower learning rate (0.0001) to avoid overfitting. The best batch size and optimizer from the first stage were employed.

Result digits dataset, Model 2 (ResNet) appears to be the top-performing model with the highest accuracy and consistently high precision, recall, and F1-scores. Model 3 (MobileNet) also performs well and can be considered as a strong alternative.

Result French dataset, Model 4 (EfficientNet) appears to be the best-performing model, exhibiting higher accuracy, precision, recall, and F1-scores. Model 3 (MobileNet) and Model 2 (ResNet) also show relatively strong performance, while Model 1 (Custom CNN) performs the poorest among the four models.

Result Arabic dataset, Model 2 (ResNet) and Model 3 (MobileNet) appear to be the top-performing models, exhibiting high accuracy, precision, recall, and F1-scores. Model 4 (EfficientNet) also performs well, while Model 1 (Custom CNN) shows lower performance compared to the other models.

General Conclusion

General conclusion:

Deafness and muteness, also known as hearing loss and speech impairment, are conditions that affect an individual's ability to hear and/or speak. Deafness refers to a partial or complete loss of hearing, while muteness refers to the inability to speak or produce speech sounds. These conditions can occur due to various factors, including genetic factors, infections, exposure to loud noise, or certain medical conditions.

Sign language is a visual-gestural language used by individuals who are deaf or have hearing loss as their primary means of communication. It relies on hand movements, facial expressions, and body language to convey meaning. Different countries and regions have their own sign languages, each with its own unique grammar, vocabulary, and syntax.

In sign language, numbers are represented using specific handshapes or gestures. The digits sign language, as mentioned, is classified into 10 classes, with each handshape representing a different number.

Arabic sign language, used primarily in Arabic-speaking countries, is classified into 28 classes, with each handshape representing a different letter.

French sign language, used in French-speaking communities, is classified into 26 classes, with each handshape representing a different letter.

In our work, we created a dataset of Arabic and French sign language images for the deaf and mute, Additionally, we obtained a separate dataset of sign language images specifically related to numbers.

And we used artificial intelligence techniques, particularly Convolutional Neural Networks, to classify the different hand signals used in communication between the deaf and mute.

After training and evaluating the models, by five CNN structures, we have recorded excellent accuracies, The best was for numbers (0.99), Arabic (0.97) then French (0.92) using the ResNet, MobileNet and EfficientNet respectively.

Bibliography

Bibliography:

[1] Dewsnup, D. (2022, November 15). Anatomy of the Ear and Understanding Hearing. Retrieved from https://www.happyearshearing.com/anatomy-of-the-ear/

[2] World Health Organization. (n.d.). Deafness and Hearing Loss. Retrieved from https://www.who.int/health-topics/hearing-loss#tab=tab_1

[3] Johns Hopkins Medicine. (n.d.). Types of Hearing Loss. Retrieved from https://www.hopkinsmedicine.org/health/conditions-and-diseases/hearing-loss/types-of-hearing-loss

[4] ‘ C. (2021, December 10). Congenital Hearing Loss: Overview and More. Retrieved from https://www.verywellhealth.com/congential-hearing-loss-5210221#citation-3

[5] Better Health Channel. (2017, April 11). Deafness - A Range of Causes. Retrieved from https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/deafness-a-range-of-causes

[6] FactDr. (2021, December 20). Muteness: Causes, Diagnosis, and Treatment. Retrieved from https://factdr.com/health-conditions/muteness/

[7] Medieval Disability Glossary. (n.d.). Muteness (Dumbness). Retrieved from https://medievaldisabilityglossary.hcommons.org/muteness-dumbness/

[8] Tietze, K. J. (2012). Clinical Skills for Pharmacists (Third Edition). In ScienceDirect. Retrieved from https://www.sciencedirect.com/topics/medicine-and-dentistry/muteness

[9] Gordon, G. N. (2023, April 25). communication. Encyclopedia Britannica. https://www.britannica.com/topic/communication

[10] Britannica, T. Editors of Encyclopaedia (2023, January 2). sign language. Encyclopedia Britannica. https://www.britannica.com/topic/sign-language

[11] Antón Dayas, I. (2019, May 28). Creation of Sign Language. National Geographic History Magazine. Retrieved from https://www.nationalgeographic.com/history/history-magazine/article/creation-of-sign-language

[12] H. Lawaniya « Computer Vision » 2020 Link: https://www.researchgate.net/publication/341313012_Computer_Vision

[13] (deep learning for computer vision) https://www.run.ai/guides/deep-learning-for-computer-vision

[14] Vonintsoa. (2021). « Classification d'image : le guide complet» . Retrieved from https://intelligence-artificielle.com/classification-d-image-guide-complet/

[15] Y. Benzaki « Introduction à l'algorithme K Nearst Neighbors (K-NN) » 2018

[16] Abdelhamid, B. (2013). "The Use of SVM Technique for Multimodal Biometric Fusion" [Master's thesis in Computer Science, University of Adrar, Algeria].

[17] Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks. In Encyclopedia of Physical Science and Technology (pp. 631-645). University of Colorado, Denver, & University of Talca, Chile. DOI: 10.1016/B0-12-227410-5/00837-1.

[18] DataScientest. Perceptron: Concept, function, and applications. [Machine Learning]. Retrieved from https://datascientest.com/en/perceptron-definition-and-use-cases.

[19] Sharma, P. (2022, March 1). Basic Introduction to Convolutional Neural Network in Deep Learning. Retrieved from https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-convolutional-neural-network-in-deep-learning/

[20] WEB site «Difference between ANN, CNN and RNN» https://www.geeksforgeeks.org/difference-between-ann-cnn-and-rnn/

[21] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi and H. Ghayvat « CNN Variants for Computer Vision : History, Architecture, Application, Challenges and Future Scope »Article Electronics 2021

[22] Choudhari, P. (2020). Understanding "convolution" operations in CNN. Medium. Retrieved from https://medium.com/analytics-vidhya/understanding-convolution-operations-in-cnn-1914045816d4

[23] Pibre, L., Chaumont, M., Ienco, D., & Pasquet, J. (2016). Étude des réseaux de neurones sur la stéganalyse. Retrieved from https://hal.science/hal-01374095v1/file/CORESA-2016-PIBRE-CHAUMONT-IENCO-PASQUET.pdf

[24] Todt, E., & Krinski, B. A. (2019). Convolutional Neural Network - CNN. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[25] B. Mohammed et B. Brahim « L`apprentissage profond (Deep Learning) pour la classification et la recherche d'images par le contenu » Professional Master's thesis UNIVERSITY KASDI MERBAH OUARGLA 2017

[26] Nepal, P. (2020, July 30). VGGNet Architecture Explained. Analytics Vidhya. Retrieved from https://medium.com/analytics-vidhya/vggnet-architecture-explained-e5c7318aa5b6

[27] "ResNet: The Basics and 3 ResNet Extensions." DataGen Tech. Retrieved from https://datagen.tech/guides/computer-vision/resnet/#:~:text=What%20Is%20Residual%20Neural%20Network,or%20thousands%20of%20convolutional%20layers.

[28] Chollet, F. (2017). Xception: Deep Learning With Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1251-1258).

[29] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Retrieved from arXiv:1704.04861

[30] (Interpreting Confusion Matrixes) https://accessibleai.dev/post/interpreting_confusion_matrixes