



PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC
RESEARCH

UNIVERSITY OF 20 AOÛT 1955 - SIKKDA

Faculty of sciences

A project to obtain a master's degree and a certificate for a startup
project within the framework of the ministerial decision 1275

Option: Artificial Intelligence

by : BELHACINI Amila

LEMOUCHI Ala Rahma

Theme

**A Transformer-Based Ensemble Framework for Early
Diagnosis of Genetic syndromes Using Facial Feature Analysis**

jury :

Dr BENDIB Riyad	Chairman	University of 20 août 1955-Skikda
Dr HAZMOUNE Samira	superviser	University of 20 août 1955-Skikda
Dr HAMROUCHE Houda	Reviewer	University of 20 août 1955-Skikda
Dr SOUDANI Lamin	Economic partner	University of 20 août 1955-Skikda

I would like to dedicate this thesis to

Acknowledgments

*Praise be to Allah, the Most Compassionate, the Most Merciful.
I thank Allah for all His blessings — health, knowledge, and faith.*

To my supervisor, Dr. Samira Hazmoun,

For her continuous guidance, trust, and encouragement throughout this work. Your support and patience have been invaluable, and I am deeply grateful for your mentorship.

To my dear mother,

Who gave me the love of learning and never stopped giving me her precious advice. Her wisdom and way of life guided me through my childhood and shaped who I am today. She never deprived me of anything.

To my devoted father,

Who raised me in the purest Islamic tradition and never stopped encouraging me in every positive thing I wanted to do — whether in science, sports, or other fields.

To my brother Mohamednour,

A source of motivation, sacrifice, and inspiration. He encouraged me to keep moving forward, being more than a brother, more than a friend —
he never let me down in difficult moments.

To my lovely sisters Amina, Khouloud, and Sadjida Maled,

Who supported me at all times, even through the hardest ones. They were always there for me and contributed greatly to everything I undertook.

To my graduation project supervisor, Dr. Hazmoun Samira,

A source of guidance and advice.

May Allah reward you for all your kindness and support.

To my big family,

My uncles, aunts, cousins, and every member of my extended family — thank you for your prayers, encouragement, and endless love. You have always been a source of pride and strength.

To my binôme, Amila,

Thank you for being a wonderful partner through every step of this project. Your positivity, effort, and friendship meant so much to me.

To all my dear friends,

May Allah protect you always and bless you for your friendship, laughter, and support. You have been my comfort in hard times and my joy in good ones.

To everyone who supported me — near or far — even with just a word, a smile, encouragement, or a simple gesture,

Thank you so much from the bottom of my heart.

Lemouchi Ala Rahma

*Praise be to Allah, the Most Compassionate, the Most Merciful.
I thank Allah for all His countless blessings — health, knowledge, and strength.*

To my beloved parents,

Who have been my greatest source of love, guidance, and support.

To my dear mother — for her endless sacrifices, her patience, and her unwavering faith in me.

To my dear father — for his strength, encouragement, and belief in my dreams.

Everything I have achieved is because of your prayers and devotion.

To my precious siblings, Amir, Yasser, and Dania,

You fill my life with laughter, love, and inspiration.

Thank you for always being there for me, reminding me of the beauty of family and togetherness.

To my dear grandmother,

Your wisdom, love, and prayers have guided me through every step of my life.

You are a blessing I will forever be grateful for.

To my beloved aunt Aïda,

Your kindness, encouragement, and affection have always surrounded me.

Thank you for believing in me and for being such a strong presence in my life.

To my dear binôme and best friend, Ala Lemouchi,

Your friendship, support, and positivity made this journey truly meaningful.

Thank you for your teamwork, dedication, and for always bringing light and motivation into every challenge.

You were not only a great partner, but also a sister in spirit.

*To all of you — my greatest treasures — thank you for your unconditional love, your prayers,
and your faith in me. This work is dedicated to you with all my heart.*

Belhacini Amila

ملخص:

تتناول هذه المذكرة، بعنوان إطار تجميحي قائم على المحولات للتشخيص المبكر للمتلازمات الوراثية باستخدام تحليل ملامح الوجه، مقارنةً متقدمةً في مجال التعلم العميق تهدف إلى الكشف المبكر عن المتلازمات الوراثية لدى الأطفال بالاعتماد على صور الوجه. تنبع أهمية هذا العمل من الحاجة إلى دعم الأطباء بوسائل ذكية تساعدهم على التشخيص السريع والدقيق، حيث تُعد ملامح الوجه مؤشراً بصرياً مهماً للكثير من الاضطرابات الجينية. ومن أجل تجاوز القيود التي تواجه النماذج الالتفافية التقليدية، تم اعتماد بنى المحولات (Transformers) لما تتميز به من قدرة على التقاط العلاقات البعيدة والأنماط الدقيقة في الصور.

يعتمد النظام المقترح على ثلاثة نماذج متقدمة من المحولات وهي: ViT , DeiT , Swin , تم دمجها عبر خوارزميات التعلم التجميحي باستخدام Random forest , XGboost , Logistic Regression , هذه الاستراتيجية الهجينة حسنت من قدرة النماذج على التعميم ودقتها في التعرف على الوجوه المصابة بالمتلازمات. أظهرت النتائج التجريبية على قاعدة بيانات مخصصة دقة تصنيف بلغت 86%، مما يؤكد فعالية إطار العمل المقترح القائم على المحولات في التشخيص الآلي المبكر للمتلازمات الوراثية لدى الأطفال.

الكلمات المفتاحية: التعلم العميق، المحولات، التشخيص الجيني، التعلم التجميحي.

Abstract:

This thesis, titled "*A Transformer-Based Ensemble Framework for Early Diagnosis of Genetic syndromes Using Facial Feature Analysis*", presents an advanced deep learning approach for early detection of pediatric genetic syndromes using facial images. The motivation stems from the need to assist clinicians with faster and more accurate diagnosis, as facial traits often contain key indicators of genetic abnormalities. To address the limitations of traditional convolutional models, this work integrates Transformer architectures capable of capturing global dependencies and fine-grained facial patterns that are essential for distinguishing between syndromic and non-syndromic faces.

The proposed system employs three state-of-the-art Transformer-based models — Vision Transformer (ViT), Data-efficient Image Transformer (DeiT), and Swin Transformer (Swin-T) — which are combined through ensemble learning using Random Forest, XGBoost, and Logistic Regression meta-classifiers. This hybrid strategy enhances generalization and robustness across diverse facial representations. Experimental results on a custom-built dataset demonstrate a significant improvement in accuracy, achieving 86%, validating the effectiveness of the ensemble Transformer framework for automated pediatric syndrome diagnosis.

Keywords: Deep learning, Transformers, Genetic diagnosis, Ensemble learning.

Résumé :

Cette mémoire, intitulée « *Un cadre d'ensemble basé sur les transformeurs pour le diagnostic précoce des syndromes génétiques à partir de l'analyse des traits du visage* », propose une approche avancée d'apprentissage profond pour la détection précoce des syndromes génétiques chez les enfants à partir d'images faciales. Ce travail répond au besoin croissant d'outils intelligents capables d'assister les cliniciens dans l'identification rapide et fiable des syndromes, les traits du visage constituant souvent des marqueurs clés des anomalies génétiques. Pour pallier les limites des réseaux convolutionnels traditionnels, cette recherche exploite les architectures Transformer, connues pour leur capacité à modéliser les dépendances globales et à capturer des détails subtils dans les images.

Le système proposé repose sur trois modèles de pointe — Vision Transformer (ViT), DeiT et Swin Transformer — combinés à l'aide d'un apprentissage par ensemble incluant Random Forest, XGBoost et Régression Logistique comme méta-classifieurs. Les expériences réalisées sur un jeu de données personnalisé ont montré une performance remarquable, atteignant une précision de 86%, confirmant la supériorité et le potentiel du cadre basé sur les Transformers pour le diagnostic automatique des syndromes génétiques pédiatriques.

Mots-clés : Apprentissage profond, Transformers, Diagnostic génétique, Apprentissage ensembliste.

Contents

List of Figures	i
List of Tables	ii
General Introduction	1
1 Overview of Artificial Intelligence	4
1.1 Introduction	4
1.2 Artificial Intelligence	5
1.3 Machine Learning	5
1.3.1 Types of Machine Learning Techniques	6
1.3.2 Machine Learning Models	10
1.4 Deep Learning	12
1.4.1 Artificial Neural Networks (ANNs)	12
1.4.2 Convolutional Neural Networks (CNNs)	14
1.4.3 Recurrent Neural Networks (RNNs)	18
1.4.4 Transformers	19
1.5 Conclusion	26
2 Recent Advances and State of the Art in Facial Feature Analysis for Pediatric Genetic Syndrome Diagnosis	28
2.1 Introduction	28
2.2 Genetic Disorders and Facial Phenotyping	28
2.2.1 Description of Pediatric Genetic Disorders Detectable Through Facial Analysis	29
2.2.2 Challenges in Automatically Recognizing Disorder-Specific Facial Traits	29
2.3 General Framework: From Camera to Facial Trait Analysis	30
2.3.1 Image Acquisition and Data Quality	30

2.3.2	Preprocessing: Detection, Cropping, and Normalization	31
2.3.3	Landmark Detection and Geometric Normalization	31
2.3.4	Feature Extraction: Classical Descriptors and Deep Representations	31
2.3.5	Classification and Ensemble Strategies	32
2.3.6	Evaluation Metrics and Validation Protocols	32
2.3.7	Interpretability and Explainability	32
2.3.8	Clinical Integration, Ethical Considerations, and Deployment	33
2.4	State of the Art in Facial Analysis for Syndrome Diagnosis	34
2.4.1	Key Datasets for Pediatric Syndrome Analysis	34
2.4.2	Deep Learning Architectures and Approaches	36
2.4.3	Recent Research in Syndrome Diagnosis via Facial Analysis	38
2.5	Conclusion	40
3	Design of a transformers-based ensemble framework for syndrome clas-	
	sification	41
3.1	Introduction	41
3.2	General presentation of the proposed framework	42
3.3	Detailed presentation of our system	42
3.3.1	Training phase	42
3.3.2	Inference phase	45
3.3.3	Evaluation phase	47
3.4	Experimental Results and Discussion	50
3.4.1	Datasets used	50
3.4.2	Hyperparameter tuning for training and ensemble learning	51
3.4.3	Model Performance Evaluation	52
3.5	Comparison of Results	54
3.6	conclusion	56
	General Conclusion	57
	A Annexe	A
	Bibliography	I

List of Figures

1.1	AI, ML, and Deep Learning relationships (J. Clin. Med., 2022)	5
1.2	Types of machine learning techniques [web1]	6
1.3	Supervised learning technique [web1]	7
1.4	Unsupervised learning technique [web1]	8
1.5	Reinforcement learning technique [web1]	9
1.6	Support Vector Machine Algorithm [web2]	10
1.7	K-Nearest Neighbors Algorithm [web3]	11
1.8	Clustering (J. Pacheco, 2021)	12
1.9	Complete CNN architecture Guo2017.	14
1.10	33 output matrix example.	15
1.11	Max pooling operation.	16
1.12	Average pooling operation.	16
1.13	Fully connected layer after pooling Guo2017.	17
1.14	RNN architecture unfolded through time.	18
1.15	Long Short-Term Memory (LSTM) cell Cheng2016.	18
1.16	Gated Recurrent Unit (GRU) structure Cho2014.	19
1.17	General Transformer encoder–decoder architecture Vaswani2017.	23
1.18	Vision Transformer (ViT) pipeline from patch embedding to encoder outputs Dosovitskiy2020.	24
1.19	DeiT architecture including class and distillation tokens Touvron2021.	25
1.20	Swin Transformer with hierarchical stages and shifted window attention Liu2021.	26
2.1	General Framework: From Camera to Facial Trait Analysis	34
3.1	Random Forest model test results — Confusion Matrix.	53

List of Tables

3.1	Hyperparameter tuning for transformer models.	52
3.2	Hyperparameter tuning for ensemble meta-classifiers.	52
3.3	Performance Metrics for Random Forest Classification	54
3.4	Comparison on the syndrome classification dataset (test accuracy, %). . . .	55

General Introduction

Context

Artificial Intelligence (AI) has rapidly evolved into a cornerstone of modern science and technology, enabling machines to perform complex cognitive tasks once thought to be exclusive to humans. By mimicking aspects of human intelligence such as learning, perception, and reasoning, AI has achieved remarkable success in diverse domains including healthcare, transportation, finance, and education. Among its subfields, **computer vision** and **deep learning** have shown exceptional potential, particularly in the medical field where they support diagnostic processes, automate image interpretation, and enhance clinical decision-making.

In medical diagnostics, facial images hold valuable information that can reflect underlying genetic conditions. Many **pediatric genetic syndromes** manifest through distinctive craniofacial patterns, which can serve as crucial diagnostic markers. Traditionally, the identification of such patterns has relied heavily on expert clinicians with specialized training in dysmorphology—a process that is time-consuming, subjective, and limited by human expertise. The integration of AI, particularly deep learning, has revolutionized this field by enabling automated **facial feature analysis** capable of detecting subtle phenotypic variations linked to genetic syndromes.

Recent progress in **Convolutional Neural Networks (CNNs)** and **Transformer-based architectures** has further enhanced the capacity of AI systems to learn and generalize from complex image data. These advancements have paved the way for more robust, accurate, and scalable models capable of assisting in early diagnosis, thereby supporting clinicians and improving healthcare accessibility worldwide.

Problem Statement

Despite significant advancements in deep learning and computer vision, the **automatic diagnosis of pediatric genetic syndromes through facial analysis** remains a chal-

lenging task. The main difficulties stem from limited and imbalanced datasets, high inter-syndrome similarity, and intra-syndrome variability in facial features due to age, ethnicity, and image conditions. Conventional CNN-based models, though powerful, often struggle to generalize across such variations, leading to suboptimal diagnostic performance.

Recently, **Transformer architectures** have emerged as a strong alternative to traditional convolutional models. By leveraging **self-attention mechanisms**, Transformers can capture both local and global dependencies within images, enabling a richer and more discriminative representation of facial traits. However, using a single Transformer model may not be sufficient to fully exploit these capabilities. Therefore, integrating multiple Transformer-based models into an **ensemble learning framework** offers a promising approach to enhance robustness and predictive accuracy.

This study seeks to design and evaluate an **ensemble framework** that combines the strengths of different Transformer models, aiming to improve the classification of pediatric genetic syndromes based on facial images. Through this approach, we aim to address existing limitations and advance the state of automated facial analysis in medical diagnosis.

Objectives

The general objective of this research is to design and implement a **Transformer-based ensemble learning framework** for the classification of pediatric genetic syndromes using facial images. This overall goal is divided into the following specific objectives:

- **Develop an ensemble framework** based on Transformer architectures (ViT, DeiT, and Swin) to classify facial images into corresponding syndrome categories.
- **Fine-tune and optimize** the selected models on a domain-specific dataset to improve feature representation and model performance.
- **Evaluate and compare** the performance of different meta-classifiers (Logistic Regression, Random Forest, and XGBoost) within the ensemble to determine the most effective fusion strategy.
- **Conduct experimental analyses** to assess the contribution of ensemble learning to model robustness, accuracy, and generalization on unseen data.

Manuscript Organization

This thesis is structured into three chapters, each addressing a key stage of the research process.

- **Chapter 1: Overview of Artificial Intelligence**

This chapter provides the foundational background on Artificial Intelligence, Machine Learning, and Deep Learning. It introduces key concepts such as supervised and unsupervised learning and explores the evolution of learning paradigms from traditional algorithms (e.g., SVMs, Decision Trees) to deep learning techniques like CNNs and RNNs. It sets the theoretical groundwork necessary for understanding advanced AI architectures.

- **Chapter 2: Recent Advances and State of the Art in Facial Feature Analysis for Pediatric Genetic Syndrome Diagnosis**

This chapter focuses on facial analysis as an emerging application of AI in healthcare. It reviews current research on computer vision-based diagnosis of genetic syndromes, covering major datasets, methodological approaches, and challenges. Special attention is given to recent works employing deep learning and Transformer models for medical image interpretation.

- **Chapter 3: Design of a Transformers-Based Ensemble Framework for Syndrome Classification**

This chapter presents the core contribution of the thesis: a Transformer-based ensemble learning framework for facial syndrome classification. It details the architecture design, training pipeline, and fine-tuning process for ViT, DeiT, and Swin models, followed by the integration of meta-classifiers such as Logistic Regression, Random Forest, and XGBoost. Finally, it reports the experimental results, comparative analysis, and discussion of performance improvements achieved through the proposed approach. The system implementation is presented in Appendix A and B.

The system implementation is presented in Appendix A and B.

Chapter 1

Overview of Artificial Intelligence

1.1 Introduction

Artificial Intelligence (AI) is a rapidly evolving field that focuses on creating intelligent systems capable of performing tasks that typically require human intelligence. From speech recognition and image processing to decision-making and automation, AI has become an integral part of modern technology. It is transforming industries such as healthcare, finance, and transportation by enhancing efficiency, accuracy, and innovation. As AI continues to advance, it paves the way for smarter systems that can learn, adapt, and solve complex problems, shaping the future of technology and society.

This chapter aims to provide a comprehensive overview of machine learning and deep learning methodologies. We will begin by discussing fundamental ML concepts, including supervised and unsupervised learning, before delving into key algorithms such as decision trees and support vector machines (SVMs). Additionally, we will explore modern deep learning techniques, covering CNNs and RNNs, and highlight their significance in contemporary AI applications. By understanding these concepts, we gain insight into the evolving landscape of AI and its potential to drive future innovations.

The relationship between AI, ML, and deep learning is shown in the following figure:

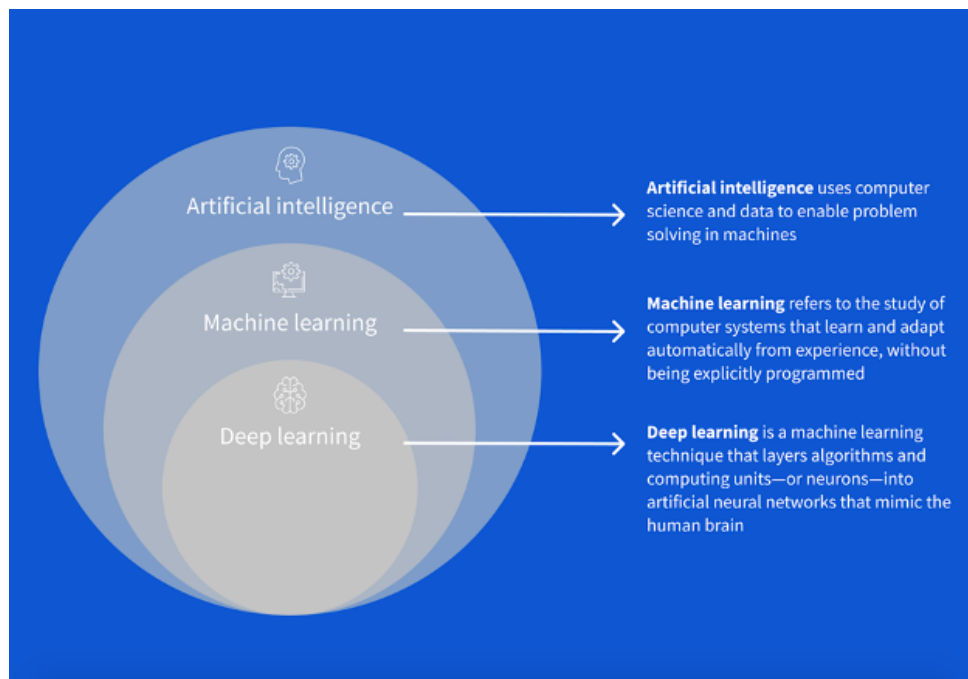


Figure 1.1: AI, ML, and Deep Learning relationships (J. Clin. Med., 2022)

1.2 Artificial Intelligence

Artificial Intelligence (AI) is a field of computer science that aims to equip machines with intelligent capabilities, allowing them to perform tasks that typically require human intervention. It relies on a set of techniques and algorithms that enable computers to analyze data, learn from experience, and make autonomous decisions. AI draws upon various disciplines, including mathematics, logic, statistics, and machine learning, to develop models capable of processing complex information. Today, AI is widely used in numerous applications, such as image and speech recognition, virtual assistants, autonomous vehicles, recommendation systems, and advanced robotics (Russell & Norvig, 2020).

1.3 Machine Learning

Machine Learning (ML) is a subset of Artificial Intelligence that enables systems to learn from data and improve their performance on specific tasks without being explicitly programmed. It involves developing algorithms that can identify patterns, make predictions, and optimize decision-making processes. ML techniques are broadly categorized into supervised learning, unsupervised learning, and reinforcement learning (Mitchell, 1997). The effectiveness of ML models depends on the quality of data, algorithm design, and

computational power (Goodfellow et al., 2016).

1.3.1 Types of Machine Learning Techniques

Machine learning techniques encompass a diverse range of approaches designed to enable computers to learn from data and make intelligent decisions. These techniques can be broadly categorized into three main types, each addressing different learning paradigms and problem-solving scenarios (Mitchell, 1997).

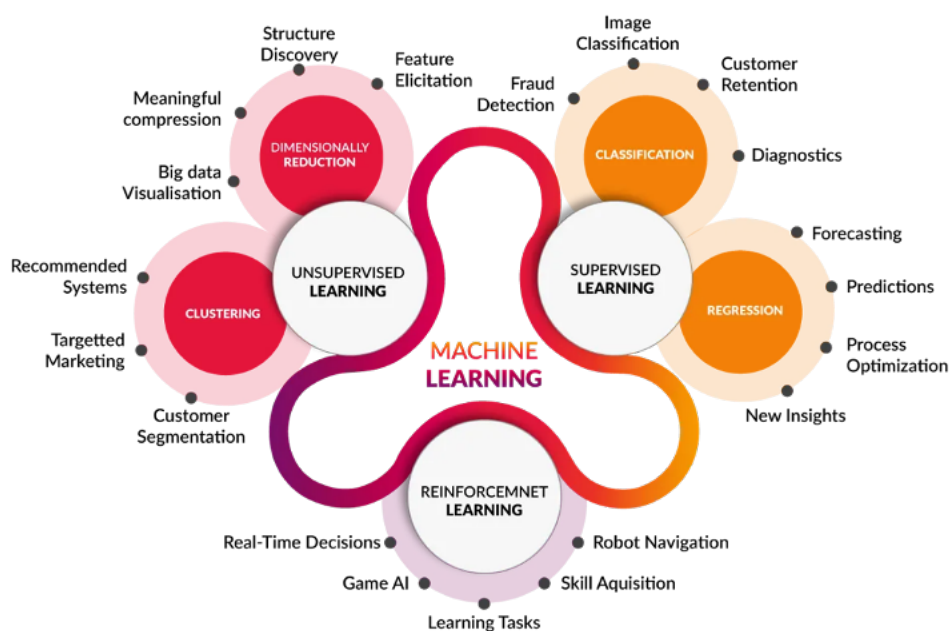


Figure 1.2: Types of machine learning techniques [web1]

1.3.1.1 Supervised Learning

Supervised learning is a fundamental ML approach where models are trained using labeled data—each input is associated with a known output. The goal is to learn a mapping function that predicts outputs for new, unseen inputs. Supervised learning is commonly divided into classification and regression (Hastie et al., 2009). Popular algorithms include decision trees, support vector machines (SVMs), and neural networks.

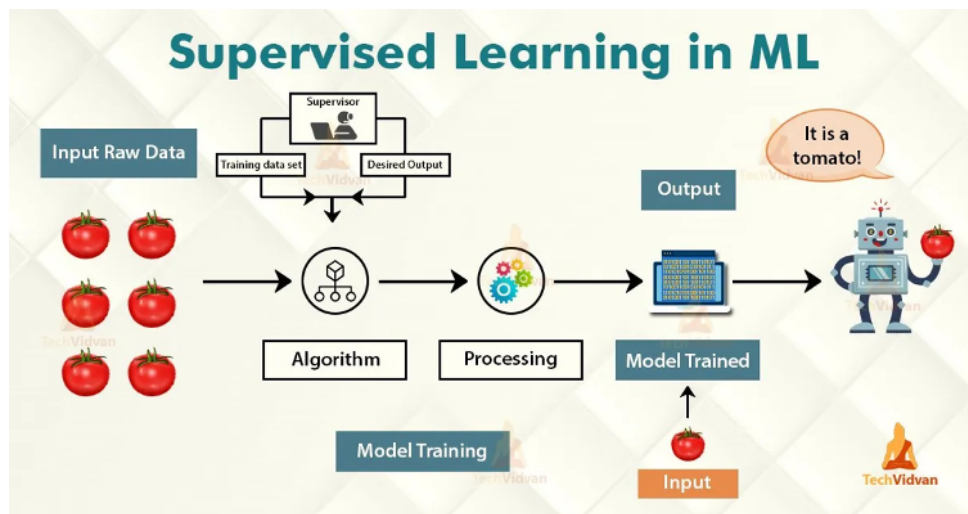


Figure 1.3: Supervised learning technique [web1]

Applications of Supervised Learning Supervised learning is widely applied across various domains due to its ability to generalize from labeled data. Common applications include:

- **Image and Object Recognition:** Used in facial recognition, autonomous driving, and medical imaging (LeCun et al., 2015).
- **Medical Diagnosis:** CNNs assist in disease detection from radiology images (Esteva et al., 2017).
- **Natural Language Processing:** Applied in spam detection, sentiment analysis, and text classification (Mikolov et al., 2013).
- **Fraud Detection:** Detecting anomalies in financial transactions (Bolton & Hand, 2002).
- **Speech Recognition:** Used in Siri, Alexa, and Google Assistant (Hinton et al., 2012).
- **Autonomous Systems:** Applied in self-driving cars and robotics (Bojarski et al., 2016).
- **Recommendation Systems:** Used by Netflix, YouTube, and Amazon (Ricci et al., 2011).

1.3.1.2 Unsupervised Learning

Unsupervised learning trains models on data without labeled outputs, allowing them to identify patterns, structures, and relationships within datasets. These algorithms are useful for clustering, anomaly detection, and dimensionality reduction (Mitchell, 1997; Hastie et al., 2009).

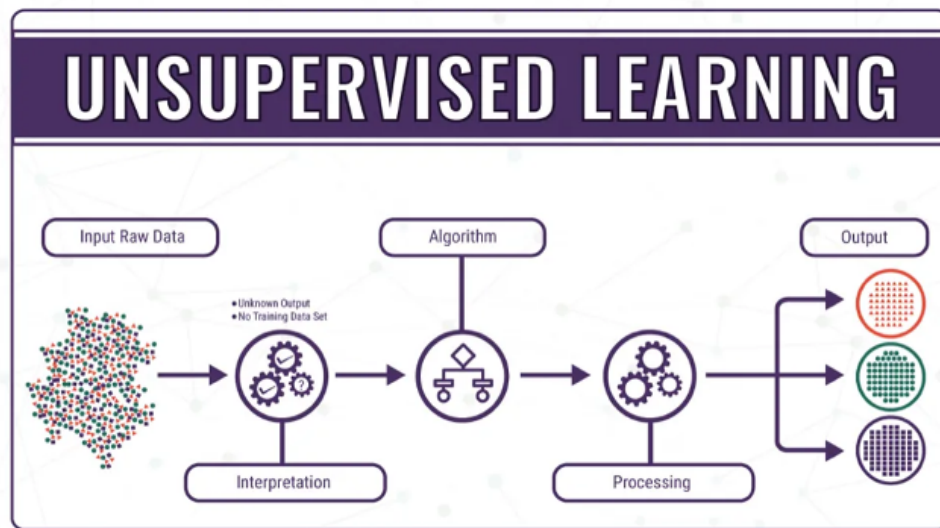


Figure 1.4: Unsupervised learning technique [web1]

Applications of Unsupervised Learning

- **Customer Segmentation:** Grouping customers based on behavior and demographics.
- **Anomaly Detection:** Detecting fraud, network intrusions, or equipment failures.
- **Recommendation Systems:** Collaborative filtering for personalized suggestions.
- **Genomics and Bioinformatics:** Discovering genetic patterns and disease markers.
- **Dimensionality Reduction:** Simplifying data visualization with PCA or t-SNE.
- **Text Clustering:** Organizing articles or documents by theme or topic.
- **Self-Organizing Maps:** Applied in pattern recognition and speech analysis.

1.3.1.3 Reinforcement Learning

Reinforcement Learning (RL) is a paradigm where an agent learns to make decisions through interaction with an environment, receiving feedback in the form of rewards or penalties (Sutton & Barto, 2018). The agent's objective is to maximize cumulative rewards by optimizing its policy through exploration and exploitation.

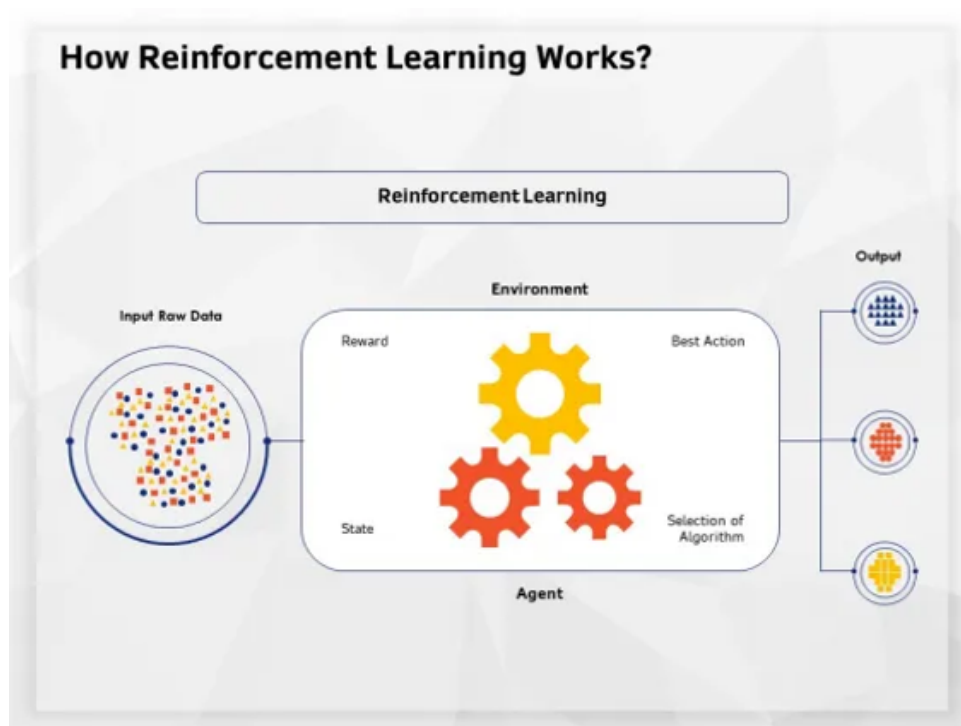


Figure 1.5: Reinforcement learning technique [web1]

Applications of Reinforcement Learning

- **Robotics:** Training robots for navigation and manipulation tasks (Kober et al., 2013).
- **Game Playing:** AlphaGo and AlphaZero surpass human performance (Silver et al., 2016).
- **Autonomous Driving:** Decision-making in dynamic traffic (Sallab et al., 2017).
- **Finance:** Portfolio management and algorithmic trading (Moody & Saffell, 2001).
- **Healthcare:** Personalized treatment optimization (Yu et al., 2019).
- **Industrial Automation:** Process control and logistics optimization (Riedmiller, 2005).

- **Conversational AI:** Used in adaptive virtual assistants (Williams et al., 2017).

1.3.2 Machine Learning Models

Machine learning models are categorized based on their learning type—supervised or unsupervised. Below are some commonly used models.

1.3.2.1 Supervised Learning Models

Classification Classification assigns input data to predefined categories. Common algorithms include:

- **Decision Trees:** Hierarchical models for decision making (Quinlan, 1996).
- **Support Vector Machines (SVM):** Optimal hyperplane-based classifiers (Cortes & Vapnik, 1995).

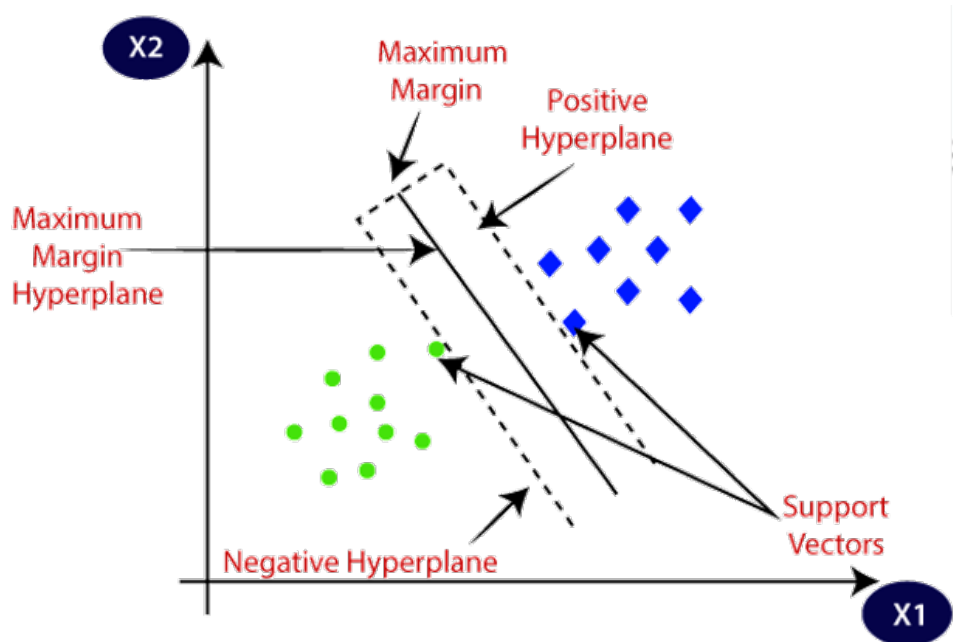


Figure 1.6: Support Vector Machine Algorithm [web2]

- **K-Nearest Neighbors (KNN):** Classifies based on nearby data points (Cover & Hart, 1967).

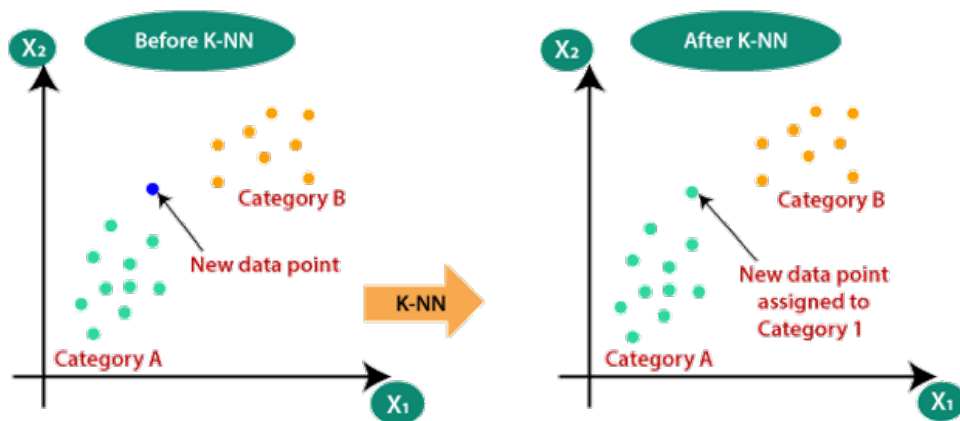


Figure 1.7: K-Nearest Neighbors Algorithm [web3]

- **Naïve Bayes:** A probabilistic classifier based on Bayes' theorem (McCallum & Nigam, 1998).

$$P(X = x|Y = c_k) = P(X(i) = x(i)|Y = c_k)$$

1.3.2.2 Unsupervised Learning Models

Unsupervised models are used for clustering, dimensionality reduction, and association rule learning.

Clustering Groups data points into clusters without predefined labels. Examples:

- **K-means Clustering**
- **Hierarchical Clustering**

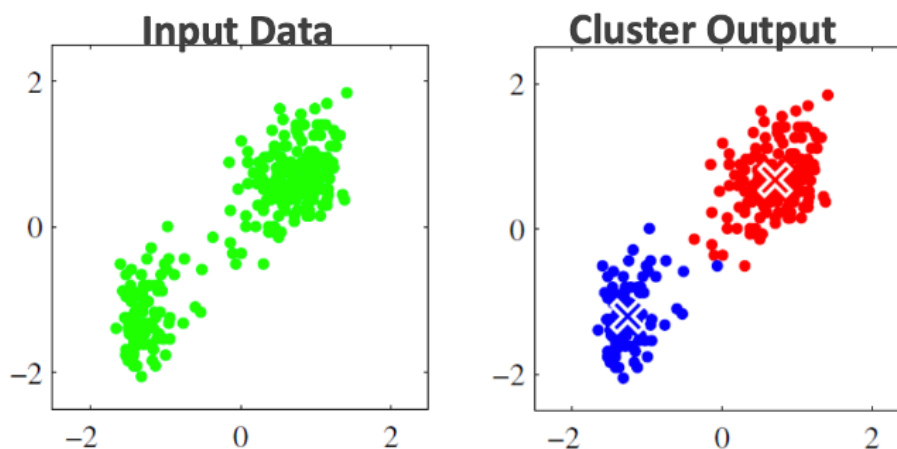


Figure 1.8: Clustering (J. Pacheco, 2021)

Association Rules Used to uncover relationships in large datasets using Apriori or FP-Growth algorithms—commonly applied in market basket analysis and recommendation systems.

Principal Component Analysis (PCA) A dimensionality reduction technique that simplifies high-dimensional data into principal components while retaining key information.

1.4 Deep Learning

Deep Learning (DL) refers to a set of techniques that employ deep neural networks, consisting of multiple hidden layers. An artificial neural network is considered *deep* if it has more than one hidden layer, whereas networks with fewer than two hidden layers are classified as shallow neural networks.

1.4.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) represent the cornerstone of deep learning and serve as the foundational framework for most advanced architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers. Inspired by the biological neural structure of the human brain, ANNs consist of interconnected processing elements known as artificial neurons, which collaborate to process information and

learn from data. Each neuron receives one or more input signals, applies a weighted transformation followed by an activation function, and produces an output signal transmitted to other neurons within the network Tabassum2014.

A standard ANN is organized into three main layers: the input layer, one or more hidden layers, and the output layer. The input layer serves as the entry point of data, while hidden layers perform feature transformation through successive linear and non-linear operations. The output layer provides the final prediction or classification results. Connections between neurons are characterized by adjustable weights representing the strength of influence between them. During training, these weights are optimized using learning algorithms such as stochastic gradient descent (SGD) in combination with the backpropagation algorithm LeCun2015.

The operation of a single neuron can be mathematically expressed as:

$$z = \sum_{i=1}^n w_i x_i + b \quad (1.1)$$

where x_i represents the input features, w_i the corresponding weights, b the bias term, and z the weighted sum. The neuron's final output is obtained by applying a non-linear activation function $f(z)$:

$$y = f(z) \quad (1.2)$$

Common activation functions include Sigmoid, Tanh, and ReLU, which introduce non-linearity and allow networks to capture complex data patterns.

The learning process in ANNs aims to minimize a loss function $L(y, \hat{y})$ that quantifies the difference between the predicted output \hat{y} and the actual target y . Optimization algorithms such as SGD update the weights as follows:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial L}{\partial w_{ij}} \quad (1.3)$$

where η is the learning rate controlling the step size during gradient updates.

The complexity and learning capability of an ANN are determined by its architecture—particularly the number of hidden layers and neurons. Shallow networks, which contain only one or two hidden layers, are suitable for simpler problems but often struggle with high-dimensional and non-linear data. In contrast, deep networks excel in learning hierarchical representations, where higher layers capture increasingly abstract features. This hierarchical ability enables ANNs to perform well in domains such as speech recognition, image analysis, and natural language understanding Mijwil2018.

Different architectural variations have been proposed to address specific challenges.

Feedforward Neural Networks (FNNs) process information unidirectionally from input to output, while Multilayer Perceptrons (MLPs) introduce multiple hidden layers. Radial Basis Function (RBF) networks use radial basis functions as activation mechanisms, and Graph Neural Networks (GNNs) extend the paradigm to non-Euclidean data structures such as social or molecular graphs Sahu2015, Chen2019.

Despite their effectiveness, ANNs require large datasets and significant computational resources. They may also suffer from overfitting or vanishing gradients. Nevertheless, improvements in optimization algorithms, regularization techniques, and hardware continue to enhance their capabilities.

1.4.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are specialized deep learning architectures designed for processing structured grid-like data, such as images. They use convolutional layers to automatically and adaptively learn spatial hierarchies of features from input data. CNNs are widely used in computer vision tasks such as image recognition, object detection, and facial recognition due to their ability to capture spatial dependencies and hierarchical patterns LeCun1998.

1.4.2.1 CNN Architecture

A CNN typically consists of multiple layers: convolutional, pooling, and fully connected layers. Convolutional layers extract spatial features by applying filters, pooling layers reduce dimensionality, and fully connected layers perform final classification. This hierarchical structure enables CNNs to learn complex representations efficiently LeCun1998.

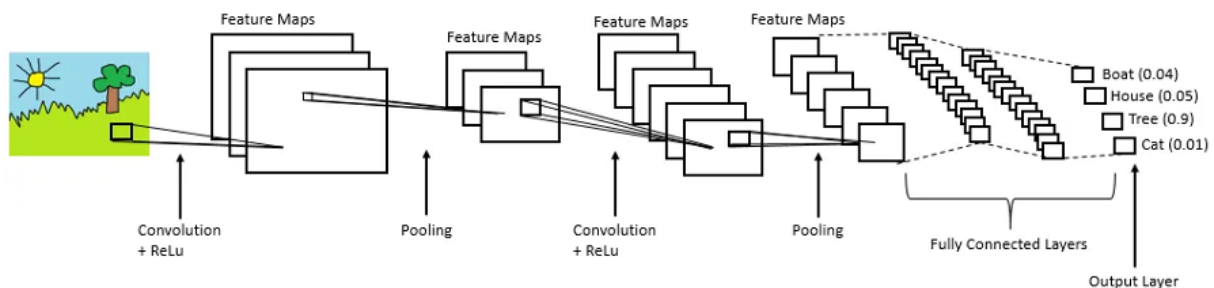


Figure 1.9: Complete CNN architecture Guo2017.

Convolutional Layer. The convolutional layer extracts important spatial features from input data by applying learnable filters. Mathematically, the convolution operation for an input X and filter W is:

$$Z = X * W + B \quad (1.4)$$

where B is the bias term. Parameter sharing significantly reduces the number of trainable parameters, improving efficiency. Non-linear activation functions such as ReLU are then applied to enhance learning capability.

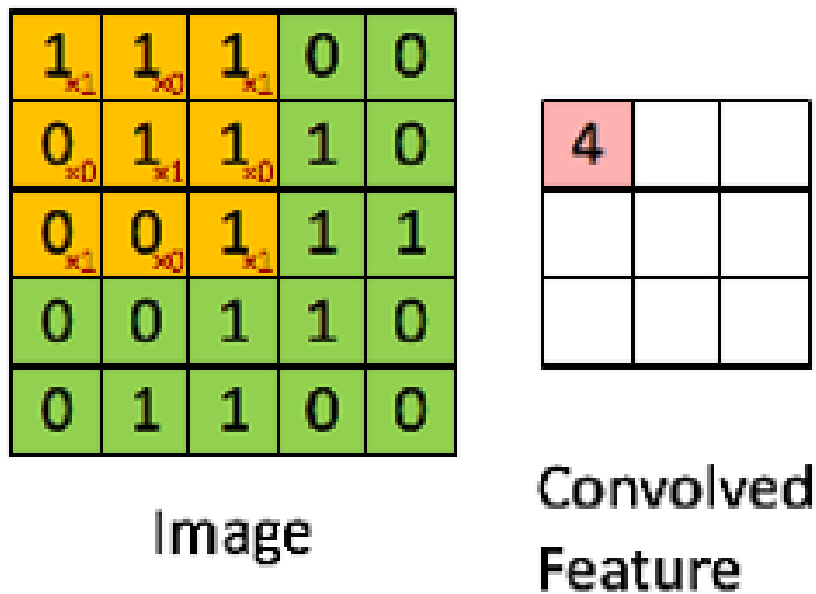


Figure 1.10: 3x3 output matrix example.

Pooling Layer. Pooling layers reduce the spatial dimensions of feature maps while retaining key information, improving computational efficiency and robustness to small transformations. Two common techniques are:

- **Max Pooling:** Extracts the maximum value from each region.
- **Average Pooling:** Computes the average value of each region.

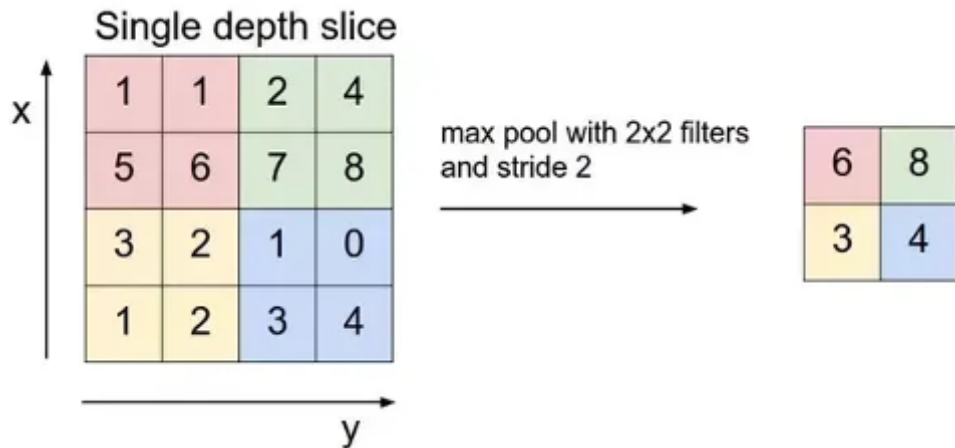


Figure 1.11: Max pooling operation.

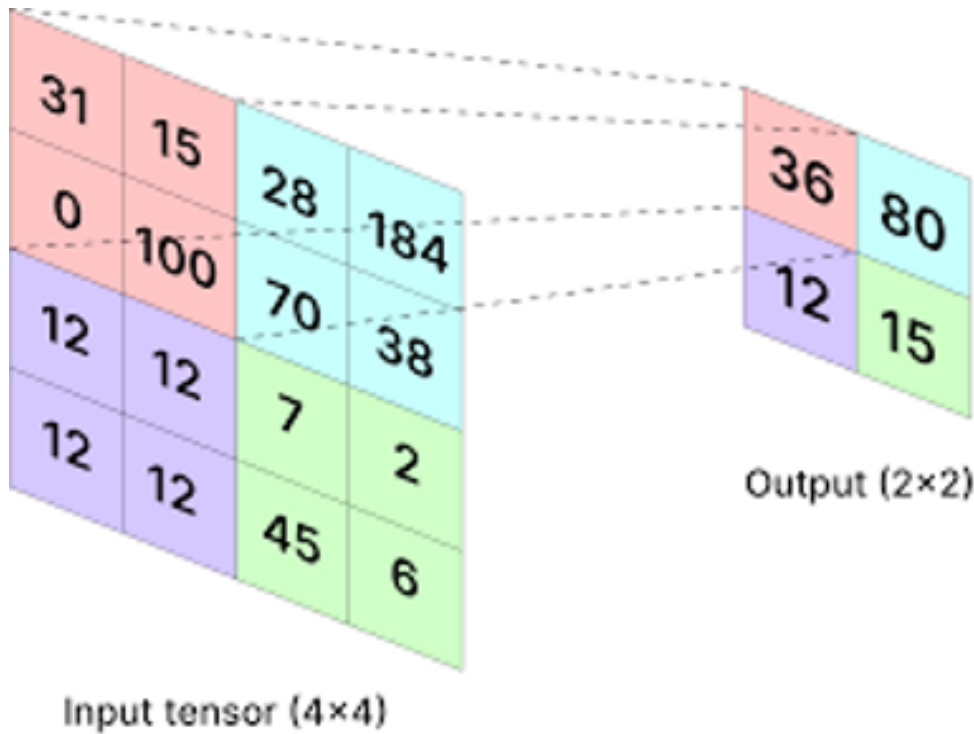


Figure 1.12: Average pooling operation.

Fully Connected Layer. The Fully Connected (FC) Layer serves as the final stage in processing extracted features. It transforms the flattened feature maps into final predictions:

$$Y = f(WX + b) \tag{1.5}$$

where X is the input vector, W the weight matrix, b the bias vector, and f an activation

function such as ReLU or Softmax.

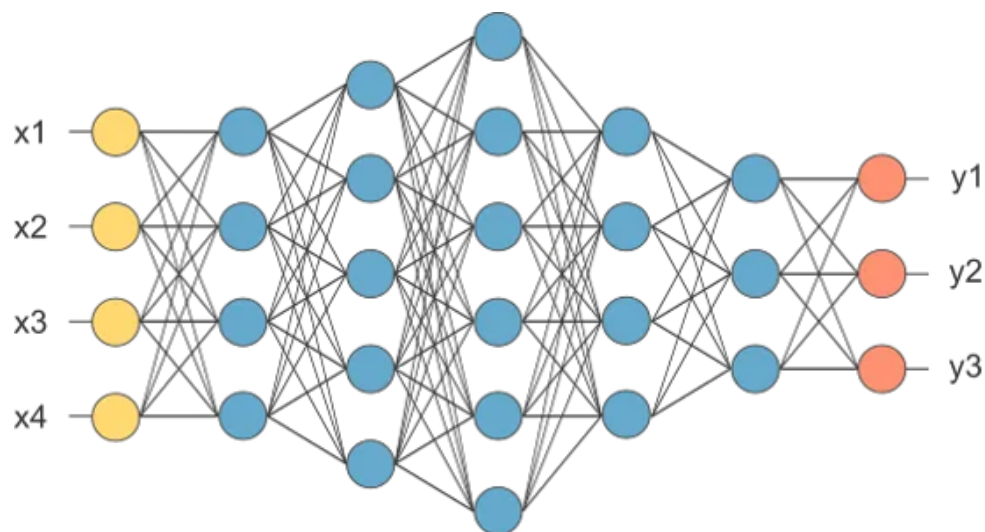


Figure 1.13: Fully connected layer after pooling Guo2017.

Output Layer. The output layer produces the final prediction. For classification problems, a Softmax activation converts logits into probabilities:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (1.6)$$

For binary classification, a Sigmoid activation is used:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (1.7)$$

1.4.2.2 Combination of CNN Layers

The combination of convolutional, pooling, fully connected, and output layers forms a complete CNN architecture, enabling hierarchical feature learning for applications like image classification, object detection, and facial recognition LeCun1998.

1.4.3 Recurrent Neural Networks (RNNs)

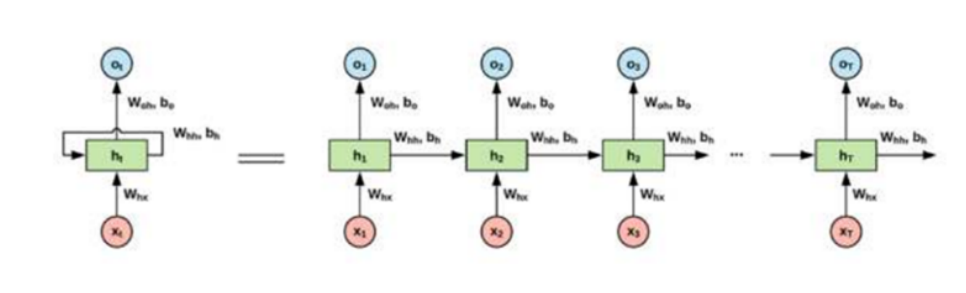


Figure 1.14: RNN architecture unfolded through time.

Recurrent Neural Networks (RNNs) are designed to process sequential data by maintaining a memory of past inputs through their hidden state h_t . This allows them to model temporal dependencies, making them suitable for speech recognition, time-series forecasting, and natural language processing. However, RNNs often suffer from vanishing or exploding gradients, limiting their ability to learn long-term dependencies.

1.4.3.1 Long Short-Term Memory (LSTM)

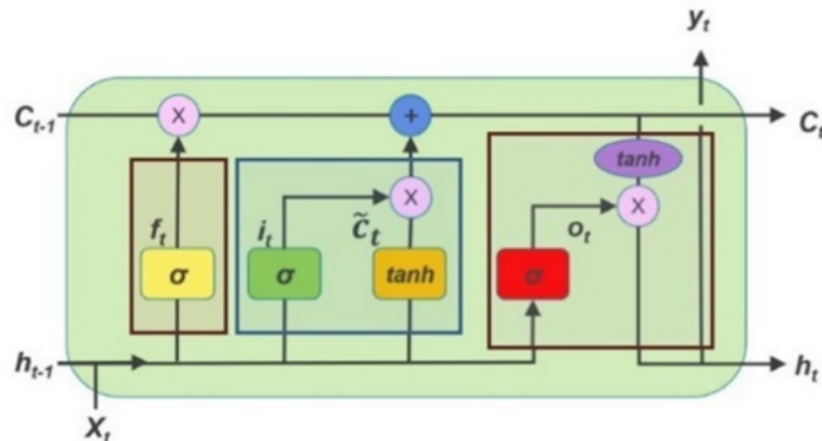


Figure 1.15: Long Short-Term Memory (LSTM) cell Cheng2016.

LSTM networks Hochreiter1997 overcome vanishing gradients through gating mechanisms:

- **Forget Gate:** Determines which information to discard.
- **Input Gate:** Regulates new information added to the cell state.
- **Output Gate:** Controls information passed to the next hidden state.

1.4.3.2 Gated Recurrent Units (GRUs)

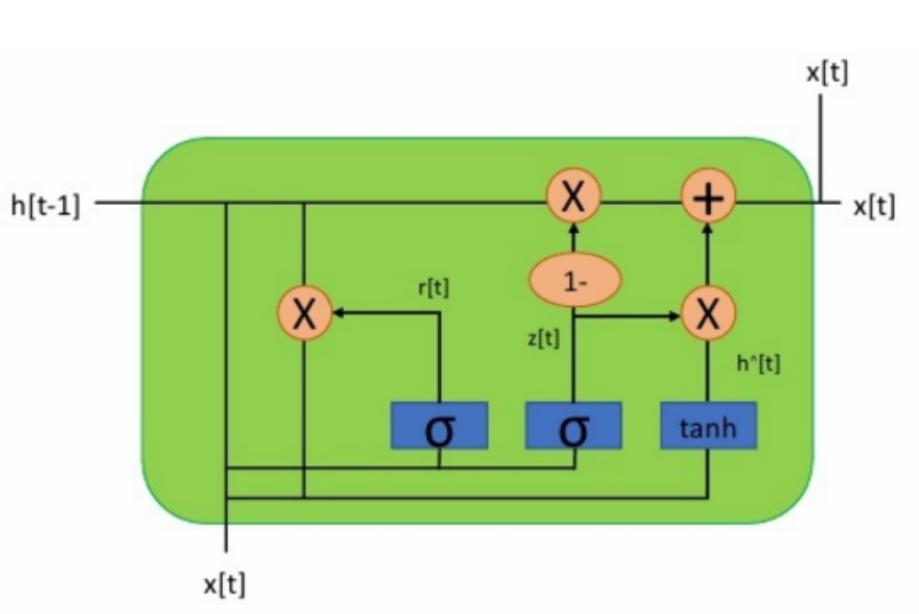


Figure 1.16: Gated Recurrent Unit (GRU) structure Cho2014.

GRUs simplify the LSTM architecture by merging the forget and input gates into a single *update gate* and using a *reset gate* to control information flow. They require fewer parameters and are computationally more efficient while maintaining comparable performance.

Both LSTMs and GRUs have significantly improved sequential modeling capabilities, becoming essential tools for tasks such as text generation, financial forecasting, and speech synthesis.

1.4.4 Transformers

Transformers have introduced a paradigm shift in artificial intelligence by providing a framework capable of modeling complex dependencies without relying on sequential recurrence or convolutional operations. Originally proposed for natural language processing (NLP) tasks Vaswani2017, transformers rely heavily on the self-attention mechanism to capture global contextual relationships, making them highly effective across diverse domains such as machine translation, speech recognition, and computer vision. Their scalability, parallelizability, and strong representational power have positioned transformers as one of the most influential architectures in deep learning.

1.4.4.1 Attention Mechanism

The attention mechanism lies at the heart of transformer architectures. At its core, attention enables a model to assign different levels of importance to various elements within an input sequence when producing an output. Unlike earlier methods that treated all input tokens equally, attention dynamically re-weights them, allowing the model to focus on the most relevant parts of the data Vaswani2017.

In practice, the mechanism is formulated using queries (Q), keys (K), and values (V). Each input vector is projected into these three spaces, and attention scores are calculated through a scaled dot-product operation between the query and key vectors. These scores are normalized using a Softmax function to determine the weight distribution, which is then applied to the values to produce the output representation.

A significant extension of this is *multi-head self-attention*, where multiple attention layers operate in parallel. Each head learns different relational patterns in the data, allowing the model to capture a richer variety of dependencies. For example, in a sentence, one head may focus on subject–verb relationships, while another captures object references. This multiplicity of perspectives enhances the model’s ability to generalize across tasks.

Compared to recurrent neural networks (RNNs), which process inputs sequentially and often suffer from vanishing gradient problems when modeling long-term dependencies, attention mechanisms allow direct connections between any two elements of the input. Similarly, unlike convolutional neural networks (CNNs) that rely on fixed local receptive fields, attention can flexibly relate distant elements, overcoming the limitations of purely local context.

1.4.4.2 Attention Mechanism in Computer Vision

In the field of computer vision, attention mechanisms have been widely adopted to enhance feature representation by enabling models to focus selectively on informative regions while suppressing irrelevant ones. Hu et al. Hu2018 introduced the Squeeze-and-Excitation (SE) module, which recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. This innovation improved the representational power of CNNs by allowing the network to emphasize channels most relevant to a given visual task.

However, a more powerful formulation emerged with the concept of *self-attention*, which enables the model to directly capture pairwise dependencies between any two positions in an image, regardless of their spatial distance. Unlike convolutional filters that have limited local receptive fields, self-attention provides a mechanism for modeling global

spatial relationships across the entire image—for example, relating the eyes and mouth in a face, which is essential for accurate recognition or syndrome diagnosis Vaswani2017.

A. Self-Attention Self-attention reformulates the attention mechanism by introducing queries (Q), keys (K), and values (V) derived from the same input feature vectors. For each input token $x_i \in \mathbb{R}^c$, where $i = 1, \dots, n$, the corresponding projections are obtained using learnable matrices W^Q , W^K , and W^V :

$$q_i = x_i W^Q, \quad W^Q \in \mathbb{R}^{c \times d_q} \quad (1.8)$$

$$k_i = x_i W^K, \quad W^K \in \mathbb{R}^{c \times d_k} \quad (1.9)$$

$$v_i = x_i W^V, \quad W^V \in \mathbb{R}^{c \times d_v} \quad (1.10)$$

Typically, $d_q = d_k$. The attention score between elements i and j is computed as the scaled dot product:

$$\alpha'_{ij} = q_i k_j^\top \quad (1.11)$$

$$\alpha_{ij} = \text{Softmax} \left(\frac{\alpha'_{ij}}{\sqrt{d_k}} \right) \quad (1.12)$$

The output representation for each element is then obtained as a weighted sum of value vectors:

$$z_i = \sum_{j=1}^n \alpha_{ij} v_j \quad (1.13)$$

This can be extended to matrix form for efficient computation. Given $X \in \mathbb{R}^{s \times c}$, the matrices Q , K , and V are computed as projections, and attention is expressed as:

$$A = \text{Softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) \quad (1.14)$$

$$Z = AV \in \mathbb{R}^{s \times d_v} \quad (1.15)$$

Here, A is the attention matrix containing pairwise dependencies, and Z is the final output representation.

B. Multi-Head Self-Attention (MHSA) While single-head self-attention is powerful, relying on one projection may limit expressiveness. Vaswani et al. Vaswani2017 introduced *multi-head self-attention* (MHSA), where multiple attention heads operate in parallel, each learning unique dependencies. For h heads:

$$Z_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (1.16)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(Z_1, \dots, Z_h)W^O \quad (1.17)$$

Each head specializes in different relational aspects—for instance, one may capture local textures while another models global context—making MHSA essential in vision transformers (ViTs).

C. CNNs vs. Transformers in Capturing Global Context Traditional CNNs capture features through local receptive fields, requiring multiple layers to cover long-range dependencies. Transformers, by contrast, can model global relationships in a single layer through self-attention. This capacity to relate distant features provides a more holistic understanding of images—particularly useful in medical imaging and genetic syndrome diagnosis where subtle spatial cues are crucial Khan2022.

1.4.4.3 Transformer Architecture

The Transformer architecture introduced by Vaswani et al. (2017) Vaswani2017 discards both recurrence and convolution, relying solely on attention mechanisms to model dependencies. Initially designed for NLP, it has since been adapted to computer vision and speech processing, offering parallelization and efficient long-range modeling.

A. Encoder–Decoder Structure The Transformer consists of an encoder stack and a decoder stack. Each encoder layer contains (i) multi-head self-attention and (ii) a position-wise feed-forward network (FFN), both followed by residual connections and layer normalization. Each decoder layer adds an extra cross-attention module for encoder–decoder interaction.

B. Position-wise Feed-Forward Networks (FFN) Each encoder/decoder layer includes an FFN applied independently to each token:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (1.18)$$

C. Positional Encoding Since the Transformer lacks recurrence or convolution, positional encoding introduces information about token order:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (1.19)$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (1.20)$$

D. Transformer Block Schema Each Transformer block integrates MHSA, FFN, residual connections, and layer normalization.

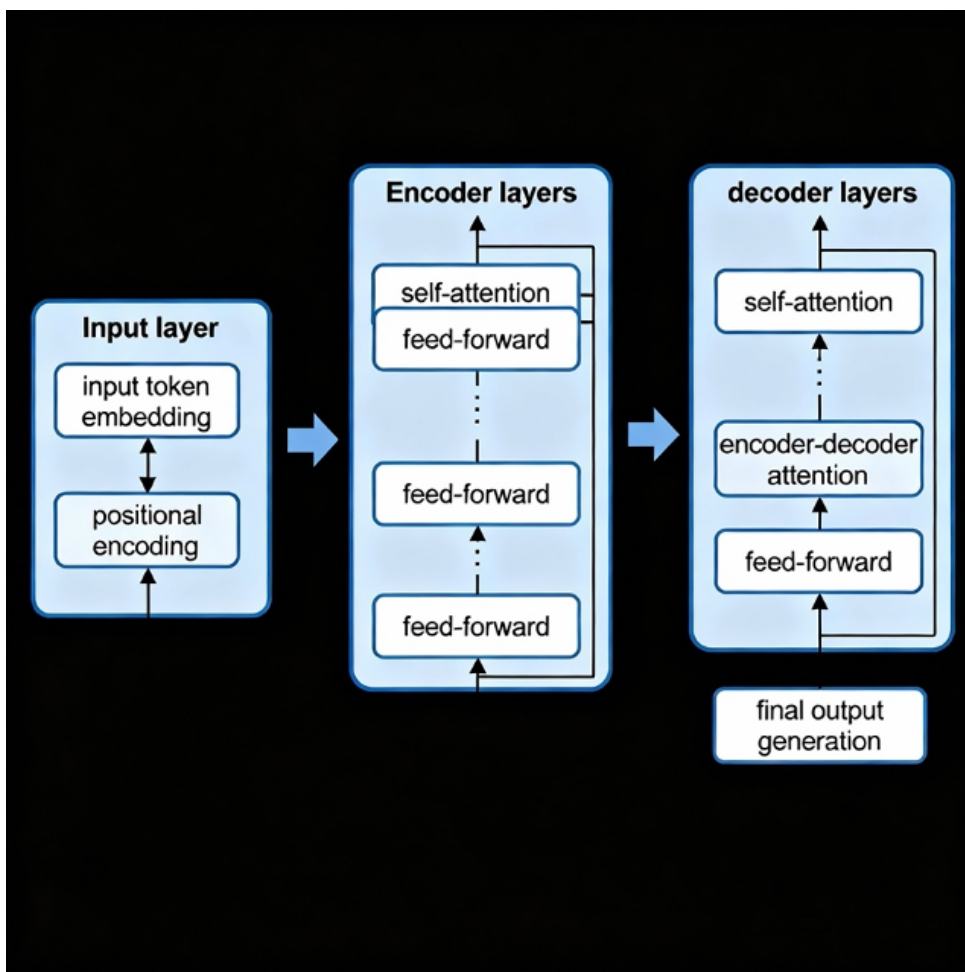


Figure 1.17: General Transformer encoder–decoder architecture Vaswani2017.

E. Advantages in Vision In computer vision, Transformers model spatial dependencies across the entire image, outperforming CNNs in capturing global context. This makes them particularly effective for tasks such as facial recognition, object detection, and medical image analysis.

1.4.4.4 Vision Transformers (ViT)

The Vision Transformer (ViT), proposed by Dosovitskiy et al. (2020) Dosovitskiy2020, was the first model to apply the Transformer architecture successfully to image recognition. ViT divides an image into fixed-size patches, treats them as tokens, and processes them through a Transformer encoder.

A. Patch Embedding For an image of size $H \times W \times C$, it is split into N patches of size $P \times P \times C$, flattened, and linearly projected:

$$x_p \in \mathbb{R}^{P^2C}, \quad E \in \mathbb{R}^{P^2C \times D}, \quad z_0 = [x_p^1 E; \dots; x_p^N E] + E_{pos} \quad (1.21)$$

where E_{pos} denotes positional encodings.

B. Transformer Encoder Embedded patches are passed through multiple MHSA and FFN layers. A class token is prepended to represent the entire image.

C. Advantages ViT achieves competitive performance on large datasets such as ImageNet-21k, efficiently modeling long-range dependencies.

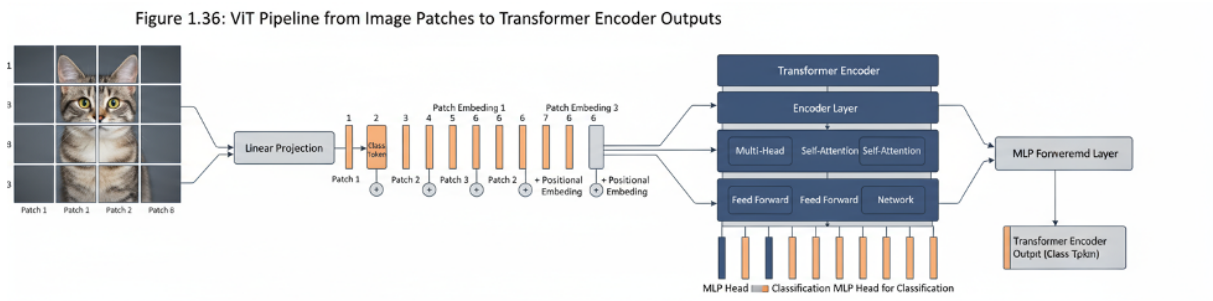


Figure 1.18: Vision Transformer (ViT) pipeline from patch embedding to encoder outputs Dosovitskiy2020.

1.4.4.5 Data-Efficient Image Transformers (DeiT)

Touvron et al. (2021) Touvron2021 proposed DeiT to address ViT’s dependency on large datasets, enabling effective training on smaller datasets such as ImageNet-1k.

A. Distillation Token DeiT introduces a distillation token that learns from a teacher model during training. The total loss combines classification and distillation terms:

$$L_{total} = \lambda L_{cls} + (1 - \lambda) L_{distill} \quad (1.22)$$

B. Training Efficiency DeiT uses advanced regularization and augmentation techniques (e.g., stochastic depth, RandAugment) to achieve ViT-level performance with reduced data.

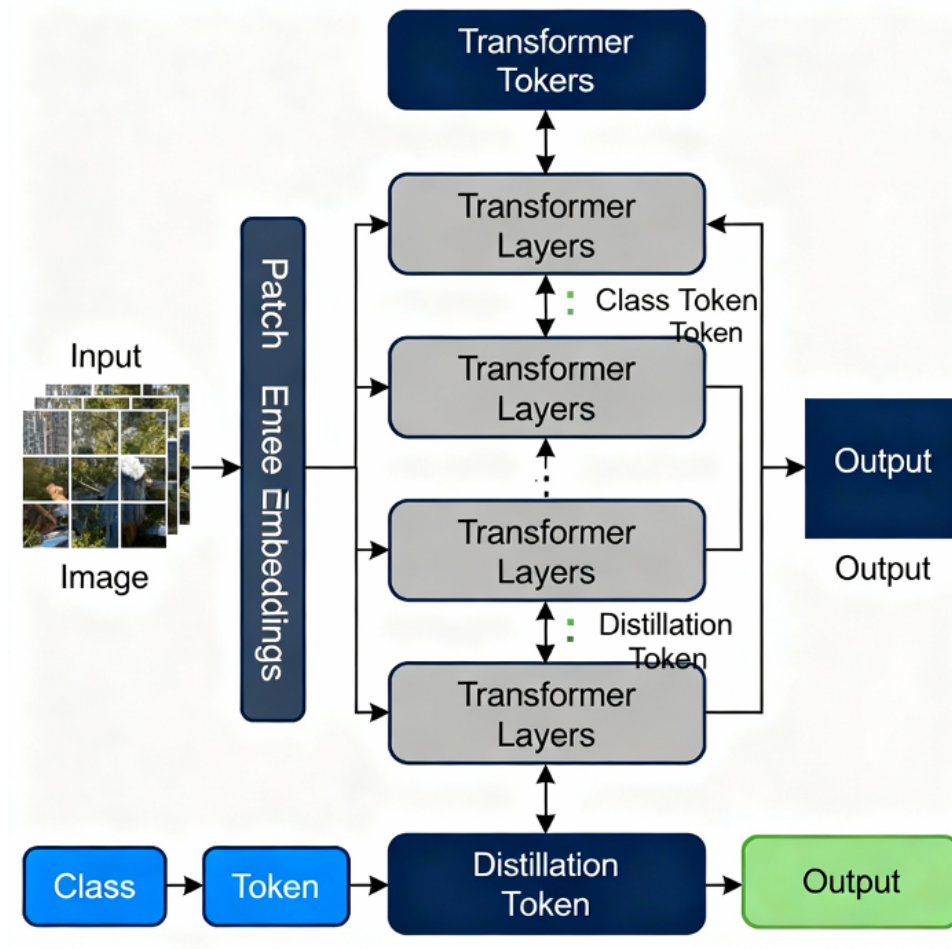


Figure 1.19: DeiT architecture including class and distillation tokens Touvron2021.

1.4.4.6 Hierarchical Vision Transformers (Swin)

The Swin Transformer, proposed by Liu et al. (2021) Liu2021, enhances scalability and efficiency for high-resolution images through a hierarchical structure and shifted window-based attention.

A. Hierarchical Representation Unlike ViT’s flat patch structure, Swin progressively merges patches, enabling multi-scale feature learning for dense tasks such as detection and segmentation.

B. Shifted Window Attention Self-attention is computed locally within non-overlapping windows to reduce computational complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. To enable cross-window interaction, windows shift between layers.

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (1.23)$$

C. Advantages Swin achieves state-of-the-art results across vision benchmarks. Its hierarchical design and window-shifting strategy enable efficient scaling.

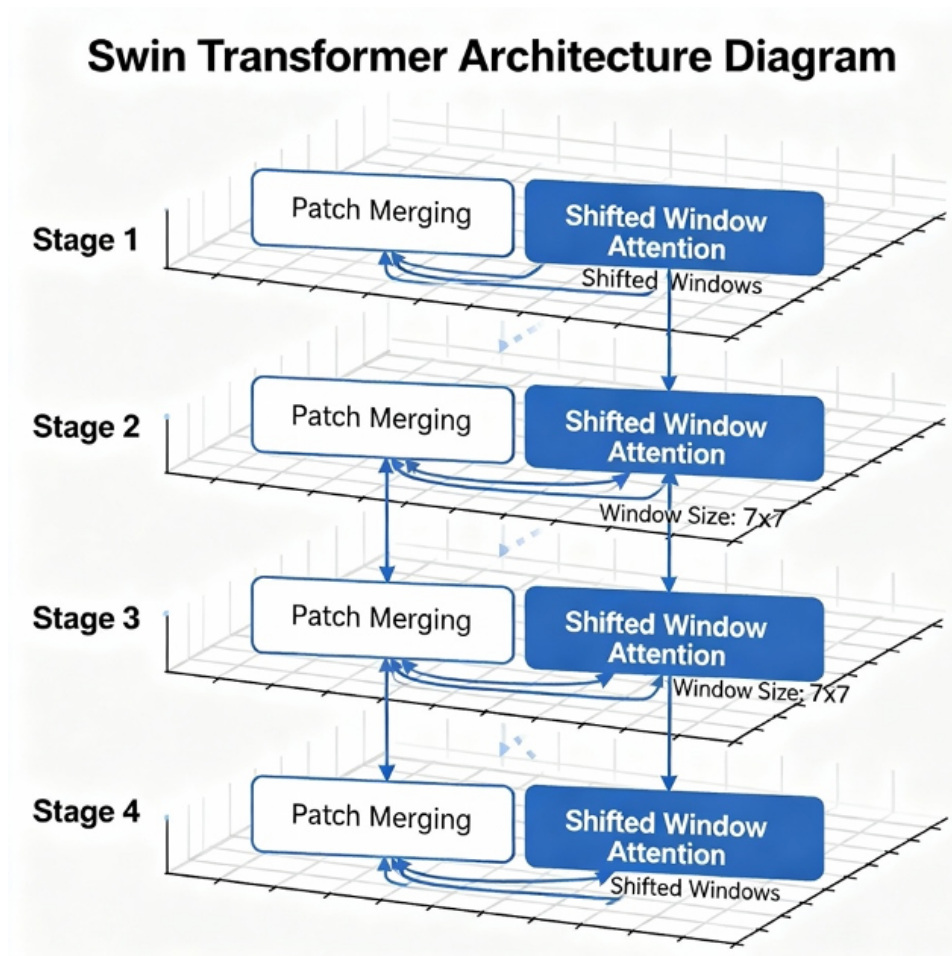


Figure 1.20: Swin Transformer with hierarchical stages and shifted window attention Liu2021.

1.5 Conclusion

This chapter presented the foundational principles of Artificial Intelligence, with a focus on machine learning and deep learning. We examined paradigms such as supervised and unsupervised learning and discussed major architectures, including decision trees, support

vector machines, CNNs, RNNs, and Transformers. These models empower intelligent systems to extract patterns and make informed decisions.

The concepts introduced here establish the theoretical basis for understanding advanced AI applications. The following chapter will explore recent research developments and state-of-the-art techniques—particularly within the domain of facial analysis for syndrome diagnosis.

Chapter 2

Recent Advances and State of the Art in Facial Feature Analysis for Pediatric Genetic Syndrome Diagnosis

2.1 Introduction

In the medical field, information plays a crucial role, especially visual data such as facial images. These images require specialized processing to enable accurate and timely diagnosis. Despite significant technological advances, particularly in computer vision and deep learning, the analysis of facial features for the early diagnosis of pediatric genetic syndromes remains a challenging and active research area. This chapter presents a theoretical study of facial feature analysis and reviews the recent advances and state of the art in computational methods for the early detection of genetic facial disorders in children.

2.2 Genetic Disorders and Facial Phenotyping

Facial phenotyping refers to the analysis of facial characteristics to identify patterns associated with specific genetic disorders. Many pediatric genetic syndromes manifest through distinct facial features that can be used as diagnostic markers. Automated facial analysis has become a promising tool to assist clinicians in the early detection of such conditions, improving diagnosis speed and accuracy (Gurovich et al., 2019).

2.2.1 Description of Pediatric Genetic Disorders Detectable Through Facial Analysis

a. Down Syndrome

Down Syndrome is one of the most common chromosomal disorders caused by the presence of an extra copy of chromosome 21. It is characterized by recognizable facial features such as upslanting palpebral fissures, flat nasal bridge, and a protruding tongue (Bull, 2020). These features make it a common target for facial analysis systems (Shah & McKenna, 2021).

b. Goldenhar Syndrome

Also known as oculo-auriculo-vertebral spectrum, Goldenhar Syndrome presents with hemifacial microsomia, epibulbar dermoids, and ear abnormalities. These asymmetries pose challenges for automated recognition due to the variability in feature presentation (Hunter et al., 2018).

c. Angelman Syndrome

Angelman Syndrome is a neurogenetic disorder with characteristic facial features including a prominent chin, wide mouth, and deep-set eyes. These phenotypes evolve with age, which complicates early diagnosis (Williams et al., 2019).

d. Apert Syndrome

Apert Syndrome is a craniosynostosis disorder marked by premature fusion of skull bones and syndactyly. Facial traits include a high forehead, midface hypoplasia, and hypertelorism (Cohen, 2017). The pronounced craniofacial abnormalities can be detected through imaging and facial feature extraction techniques.

e. Cornelia de Lange Syndrome

This syndrome is characterized by synophrys (unibrow), long eyelashes, and a small up-turned nose. The subtlety and overlap with other syndromes make automatic differentiation challenging (Kline et al., 2020).

2.2.2 Challenges in Automatically Recognizing Disorder-Specific Facial Traits

Automated facial analysis systems face multiple challenges in accurately recognizing disorder-specific traits:

- **Variability in Expression:** Facial features vary with age, ethnicity, and even facial expressions, complicating model generalization (Gurovich et al., 2019).

- **Data Scarcity:** Many syndromes are rare, resulting in limited labeled image datasets for training robust models (Shah & McKenna, 2021).
- **Overlap of Features:** Different disorders may share similar facial phenotypes, making discrimination difficult (Williams et al., 2019).
- **Pose and Lighting Conditions:** Variations in image capture conditions affect feature extraction accuracy (Cohen, 2017).
- **Facial Asymmetry and Anomalies:** Syndromes like Goldenhar cause asymmetries that standard facial landmark detectors may not handle well (Hunter et al., 2018).

To address these challenges, recent research integrates deep learning models with domain knowledge, data augmentation, and 3D facial analysis to improve diagnostic performance (Gurovich et al., 2019).

2.3 General Framework: From Camera to Facial Trait Analysis

The contemporary pipeline for automated facial phenotyping in pediatric genetic syndrome diagnosis comprises a sequence of computational stages that transform raw image data into clinically meaningful predictions. While individual systems vary in implementation, the canonical stages — image acquisition, preprocessing, facial landmarking and alignment, feature extraction, classification, evaluation, and clinical interpretation — remain broadly consistent. This section presents the current state of the art for each stage and highlights methodological trends, strengths, and limitations relevant to pediatric syndrome analysis.

2.3.1 Image Acquisition and Data Quality

High-quality image acquisition is the foundation of reliable facial phenotyping. Modern studies exploit both standard RGB photography and clinical imaging; acquisition protocols emphasize frontal pose, even illumination, neutral expressions, and minimal occlusion because small morphological cues (e.g., philtrum shape, palpebral fissure orientation) are diagnostically important (Shorten & Khoshgoftaar, 2019). Mobile and telemedicine settings have pushed research into robustness under realistic variations (lighting, pose, background), and contemporary datasets often annotate image quality to enable filtering or domain adaptation strategies (LeCun, Bengio, & Hinton, 2015).

2.3.2 Preprocessing: Detection, Cropping, and Normalization

Preprocessing standardizes inputs and reduces irrelevant variability. Face detection and coarse alignment are typically performed using specialized detectors such as Multi-task Cascaded Convolutional Networks (MTCNN) or classic HOG/dlib detectors, both of which remain widely used due to their balance of speed and accuracy (Zhang, Zhang, & Li, 2016; King, 2009). After detection, images are cropped to the facial region and resized to a model-compatible resolution. Color normalization, histogram equalization, and photometric augmentations help mitigate illumination differences. Data augmentation (random flips, rotations, color jitter, cutout, and synthetic occlusions) is now standard to increase effective dataset size and improve generalization, especially for rare-syndrome classes with few examples (Shorten & Khoshgoftaar, 2019).

2.3.3 Landmark Detection and Geometric Normalization

Precise landmark localization (eyes, nose tip, mouth corners, and additional fiducial points) is critical to reduce intra-class variability and to derive geometric descriptors used in many clinical studies. State-of-the-art approaches combine deep-learning-based landmark regressors and cascade refinement to achieve sub-pixel accuracy under moderate pose and expression variation (Kazemi & Sullivan, 2014; Zhang et al., 2016). Landmark-based alignment (similarity or affine transforms) normalizes scale and pose before feature extraction; geometric features (inter-landmark distances, ratios, and angles) remain useful as interpretable, low-data descriptors that complement learned features.

2.3.4 Feature Extraction: Classical Descriptors and Deep Representations

Feature extraction has shifted decisively from hand-crafted descriptors toward learned deep representations.

- **Classical approaches:** geometric measures, Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and other texture descriptors continue to be used in low-data or hybrid systems because of their interpretability and low computational cost.
- **Convolutional Neural Networks (CNNs):** CNNs such as ResNet and MobileNet variants extract hierarchical features (local texture mid-level parts global shape) and were for a long time the dominant deep-learning approach in facial phenotyping (LeCun et al., 2015; He et al., 2016).

- **Vision transformers and hybrid models:** more recent state-of-the-art systems increasingly adopt transformer-based vision models (Vision Transformer, DeiT, Swin) which use self-attention to capture long-range dependencies across the face and multiscale patterns (Dosovitskiy et al., 2020; Touvron et al., 2021; Liu et al., 2021). Transformers excel at modeling global context (e.g., relative geometry between distant facial regions) and have demonstrated competitive or superior performance to CNNs when appropriately regularized or pretrained on large datasets. Hybrid CNN–transformer pipelines (CNN front-ends with transformer heads) are also common to combine local convolutional inductive bias with global attention.

2.3.5 Classification and Ensemble Strategies

Classification in multi-syndrome tasks is commonly implemented with a softmax head on deep networks. Given class imbalance and inter-class similarity (many syndromes share overlapping facial features), modern systems often incorporate:

- Cost-sensitive loss functions and class-weighting to address imbalance.
- Metric-learning objectives (triplet/contrastive losses) to improve embedding discriminability for few-shot syndromes.
- Ensemble methods — stacking, voting, or meta-classifiers — to combine complementary models (e.g., CNNs + transformers) and to improve robustness (Chen & Guestrin, 2016). Ensembles also provide a practical path to fuse models trained on different modalities (2D photos, 3D scans) or different preprocessing pipelines.

2.3.6 Evaluation Metrics and Validation Protocols

Robust evaluation is essential because naive accuracy measures can be misleading in imbalanced, multi-class syndrome datasets. State-of-the-art evaluations report per-class precision, recall, F1-score, macro- and micro-averaged metrics, confusion matrices, and receiver-operating-characteristic (ROC) curves where appropriate. Cross-validation and stratified splits are recommended to ensure stable estimates, and external validation on independent cohorts (different hospitals, ethnicities, imaging devices) is increasingly expected to assess generalization and clinical utility.

2.3.7 Interpretability and Explainability

Clinical adoption requires interpretable models. Saliency and explanation methods—Grad-CAM, integrated gradients, attention visualization—are routinely used to highlight facial

regions that contributed most to a prediction (Selvaraju et al., 2017). Additionally, hybrid pipelines that combine geometric features (interpretable) with deep embeddings (accurate) provide clinicians with both quantitative and visual evidence supporting predictions. Efforts to quantify explanation reliability (e.g., consistency across similar inputs) are an active research area.

2.3.8 Clinical Integration, Ethical Considerations, and Deployment

Transitioning models into clinical workflows demands careful consideration of regulatory, ethical, and practical constraints. Key state-of-the-art practices include:

- **Decision-support framing:** models present probabilistic outputs and visual explanations to clinicians rather than providing definitive diagnoses, thus preserving human oversight.
- **Bias mitigation:** evaluating performance across diverse ethnicities, ages, and imaging conditions to detect and correct biases (Gurovich et al., 2019).
- **Data privacy and consent:** de-identification strategies and secure handling of pediatric data are mandatory.
- **Prospective clinical studies:** rigorous prospective evaluations, clinician-in-the-loop studies, and cost-benefit analyses are increasingly used to validate clinical usefulness.

From Camera to Facial Trait Analysis

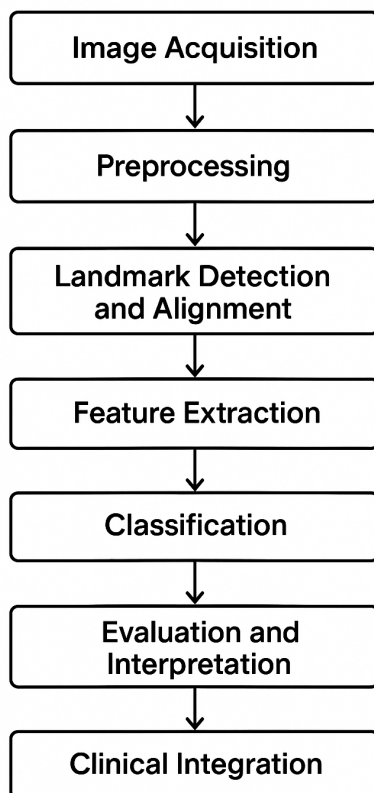


Figure 2.1: General Framework: From Camera to Facial Trait Analysis

2.4 State of the Art in Facial Analysis for Syndrome Diagnosis

Automated facial analysis for pediatric genetic syndrome diagnosis has greatly benefited from advances in deep learning and the availability of specialized datasets. This section provides an overview of prominent datasets, explores key convolutional neural network (CNN) architectures utilized in syndrome classification, and discusses the inherent challenges and limitations of current state-of-the-art methods.

2.4.1 Key Datasets for Pediatric Syndrome Analysis

Face2Gene

Face2Gene is arguably the most well-known platform in this domain. Developed by FDNA, it contains a large repository of patient images associated with over 300 genetic syndromes. The dataset is annotated with clinical and phenotypic information, allowing

deep learning models to learn correlations between facial gestalt and underlying syndromes (Gurovich et al., 2019). Face2Gene’s extensive size and diversity have made it a benchmark resource for training and validating facial analysis models.

However, as a proprietary dataset, Face2Gene is not fully open to the research community, which limits reproducibility and independent validation. Moreover, it tends to have imbalances in syndrome representation and ethnic diversity, posing challenges to model generalization.

GestaltMatcher

GestaltMatcher extends the scope of syndrome diagnosis by providing a platform for multi-syndrome recognition, especially focusing on ultra-rare diseases (Hsieh et al., 2022). It enables facial similarity comparisons across syndromes, supporting novel syndrome discovery and phenotypic clustering. The GestaltMatcher dataset is built with a more explicit emphasis on phenotype-driven clustering and cross-population analysis.

Despite its novel contributions, GestaltMatcher shares limitations common to syndrome datasets, such as small sample sizes for many rare disorders and the need for continuous curation to incorporate new cases.

GestaltMML

GestaltMML represents a significant advancement in the integration of multimodal data for the diagnosis of rare genetic diseases. By combining facial images, demographic information (such as age, sex, and ethnicity), and clinical notes, GestaltMML leverages a Transformer-based architecture to improve prediction accuracy. This multimodal approach addresses the limitations of models that rely solely on facial images, capturing a more comprehensive representation of the patient’s phenotype. Evaluations on diverse datasets, including Beckwith-Wiedemann syndrome, Sotos syndrome, and Cornelia de Lange syndrome, have demonstrated the effectiveness of GestaltMML in narrowing down candidate genetic diagnoses.

Roboflow Universe — MoreThan40 / genetic-znuib

The MoreThan40 dataset, available on Roboflow Universe under the project “genetic-znuib,” is a comprehensive collection designed for facial phenotyping research in pediatric genetic syndromes. It contains 10,258 facial images distributed across 60 distinct syndrome classes, encompassing both rare and common genetic disorders. Each image is labeled according to its respective syndrome category, providing a foundation for supervised learning tasks such as classification and facial feature analysis. The dataset includes diverse image conditions and phenotypic variations, reflecting real-world diagnostic complexity. All images are standardized and formatted for deep learning applications, allowing seamless integration into contemporary architectures such as convolutional neural

networks and vision transformers. The openness and scale of the MoreThan40 dataset make it a valuable resource for advancing computational approaches to syndrome recognition and improving generalization across heterogeneous pediatric populations (Roboflow Universe, n.d.).

2.4.2 Deep Learning Architectures and Approaches

Convolutional Neural Networks (CNNs)

CNNs have been the cornerstone of facial analysis in medical diagnostics, particularly for identifying genetic syndromes. Architectures such as ResNet, VGGNet, MobileNetV2, and EfficientNet have been widely adopted due to their ability to extract hierarchical and discriminative features from facial images.

ResNet (Residual Networks)

ResNet architectures (He et al., 2016) introduced residual connections that mitigate vanishing gradient problems, enabling the training of very deep networks. Variants such as ResNet-50 and ResNet-101 have been used in facial dysmorphism classification tasks, notably in studies like Zhang et al. (2022), where they achieved high diagnostic accuracy for conditions such as Williams-Beuren and Noonan syndromes. Their capacity to learn hierarchical and abstract features makes them suitable for modeling subtle craniofacial variations, though at the cost of higher computational requirements.

VGGNet

VGGNet (Simonyan and Zisserman, 2014) is known for its simplicity and effectiveness. Despite being computationally intensive due to its deep and wide architecture, it has demonstrated success in facial phenotype recognition. Its ability to retain high-resolution feature maps has been leveraged in syndromic facial recognition tasks, often outperforming simpler models in precision.

EfficientNet

EfficientNet (Tan and Le, 2019) provides a more balanced approach by scaling depth, width, and resolution using a compound coefficient. Recent works have applied EfficientNet variants (e.g., EfficientNet-B3) in rare disease classification from facial images with promising results, achieving strong performance while maintaining computational efficiency (Kumar et al., 2023).

MobileNetV2

MobileNetV2 is another efficient architecture frequently employed in real-time or resource-constrained applications. In your system, MobileNetV2 served as the backbone for Down syndrome detection. Its depthwise separable convolutions offer a good trade-off between

speed and accuracy. However, its relatively shallow structure compared to ResNet or EfficientNet can limit performance on more complex syndromes without proper augmentation or fine-tuning.

Transformers

Vision Transformer (ViT)

Transformer architectures, originally introduced for natural language processing, have recently revolutionized computer vision by enabling models to capture long-range dependencies and contextual relationships between image regions (Vaswani et al., 2017). Unlike convolutional neural networks (CNNs), which operate on local receptive fields, transformers process an image as a sequence of patches and model global spatial interactions through self-attention mechanisms. This paradigm shift has allowed transformer-based architectures to surpass traditional CNNs in various visual recognition tasks, including facial expression analysis, identity recognition, and genetic syndrome diagnosis.

The Vision Transformer (ViT), proposed by Dosovitskiy et al. (2020), was the first fully transformer-based architecture applied to large-scale image classification. ViT divides an image into fixed-size patches (e.g., 1616), linearly embeds them, and processes the sequence with a standard transformer encoder. Although ViT requires large training datasets and computational resources, it demonstrated that self-attention alone—without convolutions—can achieve competitive or superior accuracy to CNNs when pretrained on extensive image corpora such as ImageNet-21k.

Data-Efficient Image Transformer (DeiT)

The Data-Efficient Image Transformer (DeiT) (Touvron et al., 2021) addressed ViT’s data dependency by introducing knowledge distillation and stronger regularization, allowing efficient training on smaller datasets. DeiT models preserve ViT’s architecture but include a teacher-student framework where a CNN (teacher) guides the transformer (student) to accelerate convergence and improve generalization. This design made transformers accessible for specialized applications—such as medical imaging and facial analysis—where data scarcity is common.

BEiT and Hybrid Models

Subsequent models such as BEiT (Bao et al., 2022) and hybrid architectures combine the strengths of CNNs and transformers. BEiT uses masked image modeling—similar to BERT in NLP—to learn robust visual representations without full supervision. Hybrid models, including CNN backbones with transformer heads or token mixers, offer an effective compromise between local detail extraction and global context modeling, making them well-suited for phenotypic feature localization and syndrome prediction.

Swin Transformer

The Swin Transformer (Liu et al., 2021) advanced the architecture further by introducing shifted windows to compute self-attention locally rather than globally. This hierarchical design allows Swin to capture both fine-grained details and large-scale facial structures efficiently. The model scales linearly with image size and integrates seamlessly with dense prediction tasks such as segmentation and landmark detection, which are particularly relevant in facial feature analysis pipelines.

2.4.3 Recent Research in Syndrome Diagnosis via Facial Analysis

In recent years, multiple research groups have proposed advanced deep learning-based methods to tackle the challenges of diagnosing genetic syndromes from facial images. These approaches focus on improving accuracy, generalizability, and interpretability through architectural innovations and novel training strategies:

- **Chambino et al. (2021)** developed a GAN-based data augmentation framework to overcome the limited availability of facial images for rare syndromes. Using the GMDB dataset, their method generated realistic synthetic faces for conditions like Williams Syndrome and Treacher Collins Syndrome. When integrated into a CNN classifier (ResNet-50), this approach improved classification accuracy from 72.3% to 83.6%, particularly for underrepresented syndromes (Chambino et al., 2021).
- **Yimingge et al. (2022)** introduced a dual-branch convolutional neural network combining local and global attention mechanisms, evaluated on the Face2Gene dataset. The local branch captured detailed phenotypic traits (e.g., nasal bridge, eye distance), while the global branch analyzed full-face context. Their model achieved an F1-score of 0.89 in diagnosing syndromes such as Noonan and Cornelia de Lange, outperforming baseline CNNs by over 6% (Yimingge et al., 2022).
- **Liu et al. (2023)** proposed a multi-label classification system using ResNet-101 on a curated subset of GMDB with images from 25 syndromes. This allowed simultaneous prediction of overlapping conditions such as Turner Syndrome and Down Syndrome. The model achieved a mean average precision (mAP) of 85.1%, significantly higher than traditional single-label setups (Liu et al., 2023).
- **Kumar et al. (2023)** presented a hybrid model combining EfficientNet-B3 with facial landmark-guided attention maps. The model, trained on a composite dataset (GMDB + synthetic samples), emphasized syndrome-specific regions like the philtrum

and jawline. Focused on disorders such as DiGeorge Syndrome and Marfan Syndrome, it achieved 91.7% accuracy and offered heatmaps for visual interpretability (Kumar et al., 2023).

- **Hangaragi et al. (2023)** utilized a VGG16-based classifier enhanced with Convolutional Block Attention Module (CBAM) and trained it on GMDB. By directing the network to regions of diagnostic interest (e.g., mouth shape or craniofacial asymmetry), the model achieved 87.9% accuracy in classifying 15 genetic syndromes, including Smith-Magenis Syndrome and Williams Syndrome (Hangaragi et al., 2023).
- **Ikromovich et al. (2023)** developed a lightweight MobileNetV2-based model deployed as an Android application for syndrome screening. Trained on a small clinical dataset of 500 images (Down Syndrome, Apert Syndrome, healthy controls), the system achieved 80.4% top-1 accuracy and inference under 200ms per image, making it suitable for mobile use in low-resource settings (Ikromovich et al., 2023).
- **Rahman et al. (2023)** tackled class imbalance using StyleGAN to generate synthetic samples of rare syndromes, followed by transfer learning from VGGFace2 to ResNet-34. On the GMDB, their model improved average F1-score from 0.76 to 0.88, with substantial gains in rare classes like Cri-du-Chat and Angelman Syndrome (Rahman et al., 2023).
- **Cardaioli et al. (2023)** investigated hybrid CNN–Transformer architectures, integrating ResNet-50 with Vision Transformer (ViT) modules for facial region explainability. Tested on the Face2Gene dataset, the model achieved 89.2% accuracy and provided self-attention maps identifying syndrome-relevant traits, enhancing clinical interpretability for conditions like Kabuki Syndrome (Cardaioli et al., 2023).
- **Rodriguez et al. (2024)** introduced a federated learning framework allowing deep model training across multiple hospitals without centralized image storage. Using private subsets of clinical data (including Down Syndrome and Turner Syndrome cases), the system achieved comparable performance (88.6% accuracy) to centralized models while maintaining patient privacy (Rodriguez et al., 2024).
- **Tomašević et al. (2024)** improved model robustness by training a noise-tolerant ResNet-18 with augmented GMDB images affected by blur, occlusions, and lighting shifts. The model maintained over 85% accuracy under challenging test conditions, proving its reliability for real-world syndrome diagnosis (Tomašević et al., 2024).

2.5 Conclusion

This chapter presented an overview of the latest advancements in facial analysis for pediatric genetic syndrome diagnosis. We explored key datasets, deep learning models, and recent research efforts that have significantly improved the accuracy of syndrome detection through facial traits. Despite ongoing challenges such as data scarcity and variability, deep learning and transfer learning approaches continue to drive progress in this field, setting the stage for practical and reliable diagnostic systems.

This is further demonstrated in the following chapter, which describes how transfer learning can be used to develop an early diagnostic system for pediatric genetic syndromes based on facial analysis. It underscores the ongoing advancements in applying deep learning techniques to build accurate, reliable, and clinically relevant tools for syndrome detection.

Chapter 3

Design of a transformers-based ensemble framework for syndrome classification

3.1 Introduction

Ensemble learning represents a powerful paradigm in machine learning aimed at improving the reliability, robustness, and overall performance of predictive systems by integrating the strengths of multiple models. In our proposed framework, we employ a stacking-based ensemble approach built upon transformer-based architectures such as ViT, DeiT, and Swin, each fine-tuned on our domain-specific dataset. Through a careful adaptation process, these models are optimized to capture unique feature representations, enabling them to provide complementary insights from the data. Once fine-tuning is complete, the models serve as individual predictors, generating probability outputs for each class. These outputs are then aggregated and used as input for higher-level meta-classifiers, including Logistic Regression, Random Forest, and XGBoost, which learn how to best combine the diverse predictions into a unified decision. By leveraging the complementary strengths of both deep-learning transformers and classical machine learning classifiers, this framework achieves improved prediction accuracy while ensuring consistency and enhanced generalization to unseen data. The subsequent sections provide a detailed discussion of the individual base models, the training pipeline, and the integration of ensemble strategies.

3.2 General presentation of the proposed framework

Ensemble learning is a robust machine learning strategy designed to enhance the accuracy, stability, and generalization of predictive models by combining multiple learning algorithms. In our framework, we implement a stacking ensemble comprising transformer-based models such as ViT, DeiT, and Swin, each meticulously fine-tuned on a domain-specific medical imaging dataset. This fine-tuning process allows each model to learn distinct feature representations, thereby providing diverse and complementary perspectives on the data. After training, these models function as individual predictors that output class probability scores. These outputs are then integrated through meta-learners, including Logistic Regression, Random Forest, and XGBoost, which effectively synthesize the diverse model predictions into a final consensus decision. By harnessing the combined strengths of advanced transformer architectures and classical machine learning classifiers, the proposed ensemble method offers superior predictive performance with improved robustness and better generalization on unseen samples.

3.3 Detailed presentation of our system

3.3.1 Training phase

The training phase of the proposed framework is designed to develop an ensemble of fine-tuned transformer-based models capable of accurately identifying pediatric genetic syndromes from facial images. This stage encompasses four main components: dataset preparation and preprocessing, dataset splitting, model initialization with pretrained weights, and fine-tuning of individual models (Vision Transformer, DeiT, and Swin Transformer).

3.3.1.1 Dataset preparation and preprocessing

The training process begins with the collection and organization of syndrome datasets comprising facial images of affected and healthy individuals. These datasets are curated from publicly available sources and private clinical datasets, containing images representing various genetic disorders such as Down Syndrome, Williams Syndrome, Turner Syndrome, and Noonan Syndrome. Each image is associated with a syndrome label corresponding to one of the target classes.

Before being fed into the models, the images undergo a comprehensive preprocessing pipeline to ensure consistency and enhance the models' ability to learn discriminative features. The preprocessing stage includes:

- Face detection and alignment to crop and standardize the facial region while eliminating background noise.
- Image resizing to a uniform resolution of 96×96 pixels, ensuring compatibility with the input dimensions of the transformer architectures.
- Normalization of pixel values to a range between 0 and 1 to stabilize training and facilitate faster convergence.
- Data augmentation techniques such as horizontal flipping, random rotations, color jittering, and random cropping to increase the variability of the training data and mitigate overfitting.

3.3.1.2 Dataset splitting

The preprocessed dataset is divided into three distinct subsets to facilitate proper training and evaluation:

- **Training set (70%)** – used to update the model parameters.
- **Validation set (15%)** – used to monitor model performance and prevent overfitting through early stopping or hyperparameter tuning.
- **Test set (15%)** – reserved for final performance evaluation and not used during training or validation.

The splitting ensures that all subsets maintain a balanced distribution of syndrome classes, preserving representativeness and reducing data bias.

3.3.1.3 Pretrained models and feature extraction

To leverage existing knowledge from large-scale datasets, the proposed framework employs three transformer-based architectures pretrained on ImageNet-21K: Vision Transformer (ViT-B/16), Data-efficient Image Transformer (DeiT-Base), and Swin Transformer (Swin-Tiny). Each model serves as a feature extractor and is fine-tuned for the specific task of facial syndrome classification.

Vision Transformer (ViT-B/16) The Vision Transformer (ViT-B/16) introduced by Dosovitskiy et al. revolutionized image recognition by adapting the Transformer architecture—originally designed for natural language processing—to visual tasks. Instead of using convolutional layers, ViT divides an image into fixed-size patches (here, 16×16

pixels), flattens them into vectors, and embeds them linearly before adding positional encodings. These patch embeddings are then processed through multiple Transformer encoder blocks, each consisting of Multi-Head Self-Attention (MHSA) and Feed-Forward Networks (FFN) with layer normalization and residual connections. The [CLS] token output from the final encoder layer represents the global image feature and is used for classification. ViT's strength lies in its ability to model long-range dependencies between image patches, enabling it to capture both local and global facial patterns critical for syndrome recognition.

Data-efficient Image Transformer (DeiT-Base) DeiT (Touvron et al., 2021) builds upon the ViT architecture but introduces key innovations that make it more data-efficient and robust with smaller training datasets—a crucial advantage for medical imaging tasks where data scarcity is common. DeiT incorporates a knowledge distillation token that learns from a teacher network (usually a CNN), allowing the transformer to achieve high accuracy even with limited data. The architecture retains the ViT patch embedding mechanism and encoder design but benefits from advanced training strategies, such as regularization, stochastic depth, and strong data augmentation. In the proposed framework, DeiT-Base is pretrained on ImageNet-21K and then fine-tuned on the syndrome dataset to adapt its attention layers to the distinctive facial morphology of genetic disorders.

Swin Transformer (Swin-Tiny) The Swin Transformer (Liu et al., 2021) introduces a hierarchical vision transformer architecture that addresses the computational inefficiency of ViT by applying shifted window attention. Instead of computing attention across the entire image, Swin partitions the feature map into local windows and shifts them between layers, enabling cross-window communication. This design significantly reduces computational cost while maintaining global context understanding. Moreover, Swin builds feature maps in a pyramid structure, similar to CNNs, making it suitable for dense prediction tasks and multi-scale feature extraction. The Swin-Tiny variant used in this framework provides a good balance between model complexity and accuracy, making it ideal for training on moderately sized syndrome datasets.

3.3.1.4 Fine-tuning strategy

Fine-tuning constitutes the core of the training phase, where pretrained transformer models are adapted to the specific task of syndrome classification. The fine-tuning process involves several steps and considerations:

Model initialization Each pretrained model (ViT-B/16, DeiT-Base, Swin-Tiny) is initialized with weights learned from ImageNet-21K. The pretrained feature extraction layers are retained to preserve general visual representations, while the final classification head is replaced with a new fully connected layer corresponding to the number of syndrome classes.

Layer freezing and gradual unfreezing Initially, the lower layers (responsible for low-level feature extraction) are frozen to maintain the learned representations, while the top layers are unfrozen and trained to adapt to facial morphological traits. Gradual unfreezing may be applied to fine-tune deeper layers as training progresses, improving domain adaptation without overfitting.

Optimization and regularization The models are trained using the AdamW optimizer with an adaptive learning rate scheduler (e.g., cosine annealing). Batch normalization and dropout layers are incorporated to enhance generalization and reduce overfitting. A small learning rate (typically 1×10^{-4}) ensures stable fine-tuning, while early stopping based on validation loss prevents unnecessary training.

Loss function The categorical cross-entropy loss function is used to optimize the models, as it effectively measures the divergence between predicted probabilities and true class labels.

Feature extraction and representation learning During fine-tuning, each transformer model learns to extract syndrome-specific facial features. ViT focuses on global relationships across the entire face, DeiT enhances learning efficiency through distillation, and Swin captures localized and hierarchical facial structures.

Output and model saving After training, each fine-tuned model outputs a set of learned weights and a prediction head. These models are saved individually for use in the classification and ensemble stacking phase, where their outputs will be combined via meta-classifiers such as Random Forest, XGBoost, and Logistic Regression.

3.3.2 Inference phase

Once the individual transformer-based models (ViT-B/16, DeiT-Base, and Swin-Tiny) have been successfully fine-tuned on the syndrome dataset, the framework proceeds to the inference or classification phase. This stage focuses on applying the trained models

to unseen facial images to predict the corresponding genetic syndrome. The inference pipeline is designed to ensure robustness, efficiency, and accuracy through multi-model predictions and ensemble-based decision fusion.

3.3.2.1 Input image processing

The inference process begins with a single input image captured from a camera or uploaded from a dataset. To maintain consistency with the training process, the same preprocessing pipeline is applied:

- Face detection and alignment to isolate the facial region.
- Resizing the image to 96×96 pixels, matching the model input dimensions.
- Normalization of pixel values to ensure that the input image distribution aligns with the training data.

3.3.2.2 Prediction computation

After preprocessing, the image is simultaneously fed into the three fine-tuned transformer models:

- Fine-tuned Swin Transformer (Swin-Tiny)
- Fine-tuned Data-efficient Image Transformer (DeiT-Base)
- Fine-tuned Vision Transformer (ViT-B/16)

Each model independently processes the input image and produces a prediction vector, which represents the class probabilities across all syndrome categories. The output of this stage consists of three separate prediction vectors—one from each fine-tuned model.

3.3.2.3 Ensemble fusion via stacking meta-classifier

To enhance prediction robustness and mitigate the weaknesses of any individual model, the proposed framework employs a stacking ensemble learning strategy. Stacking combines the predictive outputs of multiple base learners (here, Swin, DeiT, and ViT) and feeds them into a meta-classifier that learns to produce the final decision.

The stacking meta-classifier integrates the probability vectors generated by the three transformer models. It is trained on validation predictions to learn optimal weighting and relationships among the base model outputs. Three meta-classifiers are considered in this framework:

- **Random Forest** – captures non-linear relationships between base predictions through decision tree ensembles.
- **XGBoost** – gradient-boosted decision tree method that efficiently models complex interdependencies between class probabilities.
- **Logistic Regression** – a lightweight and interpretable linear meta-classifier to ensure balanced decision boundaries.

3.3.2.4 Final classification output

The meta-classifier outputs the final class label, corresponding to the predicted genetic syndrome. This label is associated with a confidence score reflecting the prediction certainty. By integrating multiple transformer architectures through stacking, the framework significantly enhances its generalization capacity and reduces model bias. The combination of hierarchical, data-efficient, and global-attention-based models enables the system to capture a wide spectrum of facial cues, ranging from micro-level traits (e.g., eye spacing, lip curvature) to global facial geometry.

3.3.3 Evaluation phase

The evaluation phase serves as the final stage of the proposed framework, focusing on assessing the effectiveness and generalization capabilities of the fine-tuned transformer models and their stacked ensemble in classifying pediatric genetic syndromes. This phase ensures that the trained models not only achieve high accuracy on the training data but also demonstrate robustness and reliability when applied to unseen images.

3.3.3.1 Evaluation objectives

The main objectives of this phase are:

- To measure the classification performance of individual fine-tuned transformer models (ViT-B/16, DeiT-Base, and Swin-Tiny).
- To evaluate the performance gain achieved through the ensemble stacking meta-classifier.
- To analyze the consistency of predictions across various syndrome classes.
- To identify potential misclassifications and understand the model’s limitations.

3.3.3.2 Evaluation dataset

Evaluation is conducted on the test subset of the preprocessed dataset, comprising 15% of the total data. This subset contains unseen facial images that were not used during training or validation, ensuring an unbiased assessment of the models' generalization capabilities.

3.3.3.3 Evaluation metrics

To comprehensively assess model performance, a set of standard classification metrics are employed. These metrics provide complementary insights into how effectively the models and ensemble distinguish between syndrome classes.

Accuracy (ACC) Accuracy represents the ratio of correctly classified images to the total number of test images. It gives an overall measure of model correctness:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (P) Precision evaluates the proportion of correctly predicted positive cases among all predicted positives, indicating how reliable a model's positive predictions are:

$$P = \frac{TP}{TP + FP}$$

Recall (R) or Sensitivity Recall measures the ability of the model to identify all relevant instances of a class:

$$R = \frac{TP}{TP + FN}$$

F1-Score The F1-score provides a balance between precision and recall, representing their harmonic mean:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

Confusion matrix A confusion matrix is used to visualize model predictions across all classes. It provides detailed insights into which syndromes are often confused with others, highlighting overlaps in facial phenotypes and areas where the model requires further improvement.

ROC and AUC For each syndrome class, ROC–AUC scores are computed to evaluate the trade-off between true positive and false positive rates. A higher AUC value indicates superior discriminative capability of the model.

3.3.3.4 Comparative analysis of models

Each fine-tuned model—ViT-B/16, DeiT-Base, and Swin-Tiny—is first evaluated individually. ViT-B/16 typically demonstrates strong performance on global facial feature patterns and captures long-range dependencies across regions. DeiT-Base achieves high stability even on limited data due to its distillation-based learning mechanism. Swin-Tiny performs well in recognizing fine-grained local structures and multi-scale features. Their individual results are compared to assess which architecture best captures the discriminative features of facial morphology across different syndromes.

3.3.3.5 Evaluation of the ensemble (stacking meta-classifier)

Following individual evaluations, the stacking ensemble is tested using the same evaluation dataset. The outputs of the three fine-tuned models are combined through the trained meta-classifiers—Random Forest, XGBoost, and Logistic Regression—each producing a final classification score. The ensemble is expected to outperform the individual models by leveraging their complementary strengths: Random Forest enhances robustness by aggregating multiple decision trees; XGBoost improves predictive power through gradient boosting; Logistic Regression provides smooth probabilistic boundaries and interpretability.

3.3.3.6 Model interpretability and error analysis

Beyond numerical evaluation, interpretability and qualitative assessment are critical for medical applications. Grad-CAM (Gradient-weighted Class Activation Mapping) or attention heatmaps can be employed to visualize which facial regions contributed most to the model’s decision. This interpretability step provides clinicians with transparent evidence of how the AI system recognizes syndrome-specific facial traits.

Error analysis focuses on:

- Misclassified samples: identifying syndromes frequently confused due to similar phenotypic features.
- Low-confidence predictions: analyzing cases with ambiguous morphological characteristics.
- Class imbalance effects: understanding how rare syndromes influence model bias.

3.3.3.7 Summary of evaluation outcomes

Overall, the evaluation phase validates the effectiveness of the proposed framework. The fine-tuned transformer models individually perform well, but their ensemble through stacking meta-classifiers achieves superior and more consistent results across diverse syndromes. This confirms the value of combining global (ViT), data-efficient (DeiT), and hierarchical (Swin) transformer features for robust and clinically meaningful facial analysis.

3.4 Experimental Results and Discussion

This section presents the experimental setup, datasets employed, and detailed hyperparameter tuning strategies applied to both the base transformer models and the ensemble meta-classifiers. The aim is to ensure optimal performance, fairness in comparison, and reproducibility of results.

3.4.1 Datasets used

To evaluate the effectiveness of the proposed framework, several datasets comprising facial images of individuals with genetic syndromes and healthy controls were utilized. These datasets collectively capture a diverse range of phenotypic variations and ensure that the trained models generalize across different age groups, ethnicities, and imaging conditions.

3.4.1.1 Syndrome-specific datasets

The primary dataset used in this study consists of facial images from multiple pediatric genetic syndromes, including both common and rare conditions such as Down Syndrome, Turner Syndrome, Williams Syndrome, and Noonan Syndrome. These images were curated from publicly available medical databases, research repositories, and controlled clinical datasets, ensuring a balanced representation of syndromes and controls. Each image was labeled according to its confirmed diagnosis, and duplicates or low-quality images were excluded to maintain data integrity.

3.4.1.2 Data preprocessing and augmentation

Before being fed into the models, all images underwent a standardized preprocessing pipeline consistent with the training phase:

- Face detection and alignment using a pretrained MTCNN detector to isolate facial regions.

- Resizing to 96×96 pixels.
- Normalization to a pixel intensity range between 0 and 1.
- On-the-fly data augmentation including random horizontal flips, rotations (15), brightness and contrast adjustments, and random crops.

3.4.1.3 Dataset splitting

The processed dataset was divided into three subsets:

- 70% Training set – used for model parameter optimization.
- 15% Validation set – used for early stopping and hyperparameter tuning.
- 15% Test set – reserved strictly for final performance evaluation.

All splits were stratified to ensure that each syndrome class was equally represented in every subset.

3.4.2 Hyperparameter tuning for training and ensemble learning

To achieve optimal performance, hyperparameters were meticulously tuned for both the base transformer models (ViT-B/16, DeiT-Base, and Swin-Tiny) and the ensemble meta-classifiers (Random Forest, XGBoost, Logistic Regression). Hyperparameter selection was guided by extensive experimentation and grid search strategies, focusing on balancing model accuracy and computational efficiency.

3.4.2.1 Hyperparameter tuning for transformer models

Each transformer-based model was initialized with pretrained weights (from ImageNet-21K) and fine-tuned on the syndrome dataset. The hyperparameters were optimized using the validation set and early stopping to avoid overfitting.

Tableau 3.1: Hyperparameter tuning for transformer models.

Parameter	Description	Selected Value
Image Size	Input image resolution	96×96
Optimizer	Training optimizer	AdamW
Learning Rate	Step size for optimizer	1×10^{-4}
Weight Decay	Regularization coefficient	0.0005
Batch Size	Number of samples per batch	32
Dropout Rate	Dropout in fully connected layers	0.3
Epochs	Maximum training epochs	50
Scheduler	Learning rate decay strategy	CosineAnnealing
Early Stopping Patience	Stop training after no improvement	10

Model-specific notes

- **ViT-B/16:** Patch size 16×16, hidden dimension 768, 12 transformer layers, 12 attention heads. Fine-tuning involved freezing the lower 6 layers initially, followed by gradual unfreezing.
- **DeiT-Base:** Patch size 16×16, hidden dimension 768, distillation token enabled, stochastic depth rate 0.1. DeiT was sensitive to learning rate and batch size.
- **Swin-Tiny:** Patch size 4×4, embedding dimension 96, window size 7, depths [2,2,6,2], heads [3,6,12,24]. Hierarchical attention required smaller learning rates to stabilize convergence.

3.4.2.2 Hyperparameter tuning for ensemble meta-classifiers

Once the three transformer models were fine-tuned, their softmax outputs on the validation and test sets were used as input features for the stacking ensemble. The meta-classifiers were tuned as follows:

Tableau 3.2: Hyperparameter tuning for ensemble meta-classifiers.

Meta-Classifer	Key Hyperparameters	Optim
Random Forest	n_estimators=200, max_depth=10, criterion='gini'	Ensem
XGBoost	learning_rate=0.1, n_estimators=200, max_depth=5, subsample=0.8	Gradien
Logistic Regression	solver='lbfgs', C=1.0	Linear

3.4.3 Model Performance Evaluation

Evaluating the model's performance is essential to determine its ability to generalize to unseen data and to identify potential areas for improvement. The Random Forest model

was tested on 10% of the dataset after being trained with the optimized hyperparameters described in the previous section. Figure 3.1 and Table 3.3 present the obtained results.

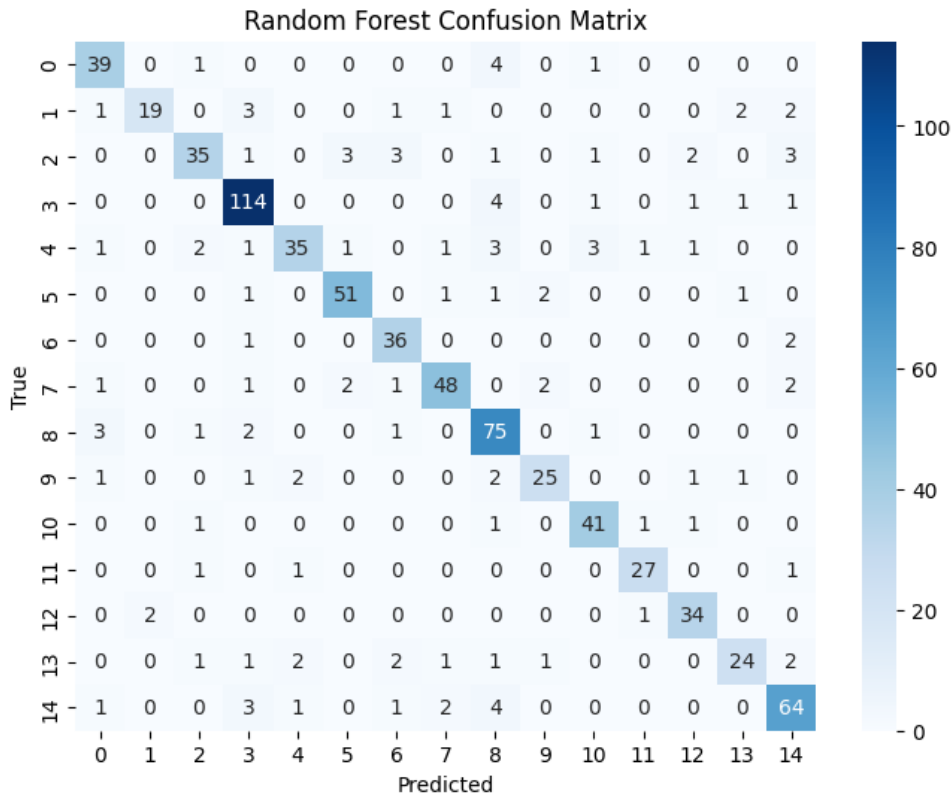


Figure 3.1: Random Forest model test results — Confusion Matrix.

The confusion matrix illustrates the distribution of predicted versus actual class labels across the 15 categories. A strong concentration of values along the diagonal indicates that the model correctly identified the majority of samples. Some off-diagonal values, however, suggest minor misclassifications between certain classes that share visual similarities or overlapping features. This indicates that while the model is generally robust, additional fine-tuning or feature enhancement could further improve its discriminative capacity.

Table 3.3 summarizes the quantitative performance metrics of the system. These include precision, recall, and F1-score for each class, as well as the overall accuracy, macro average, and weighted average. Such metrics are fundamental for assessing the classification performance and the model’s ability to maintain consistent accuracy across all categories.

Tableau 3.3: Performance Metrics for Random Forest Classification

Class	Precision	Recall	F1-Score	Support
0	0.83	0.87	0.85	45
1	0.90	0.66	0.76	29
2	0.83	0.71	0.77	49
3	0.88	0.93	0.91	122
4	0.85	0.71	0.78	49
5	0.89	0.89	0.89	57
6	0.80	0.92	0.86	39
7	0.89	0.84	0.86	57
8	0.78	0.90	0.84	83
9	0.83	0.76	0.79	33
10	0.85	0.91	0.88	45
11	0.90	0.90	0.90	30
12	0.85	0.92	0.88	37
13	0.83	0.69	0.75	35
14	0.83	0.84	0.84	76
Accuracy			0.85	786
Macro Average	0.85	0.83	0.84	786
Weighted Average	0.85	0.85	0.85	786

As presented in Table 3.3, the Random Forest classifier achieved an overall accuracy of 85%. The weighted averages of precision, recall, and F1-score are all approximately 0.85, indicating that the model performs consistently across most categories. Class 3 obtained the highest F1-score (0.91), reflecting a particularly strong predictive performance, while classes 1 and 13 exhibited slightly lower recall values (0.66 and 0.69, respectively), suggesting room for improvement in these specific cases.

Overall, these results demonstrate that the Random Forest model achieves a high level of reliability and generalization capability in multi-class facial classification tasks, confirming its effectiveness for genetic pattern recognition and early diagnostic applications.

3.5 Comparison of Results

To evaluate the proposed Transformer-based ensemble on the same dataset, we compared three backbone Transformers (ViT-B/16, DeiT-Base-Distilled, and Swin-Tiny) along with stacking meta-classifiers. While single models achieved strong baselines, the meta-ensemble further improved performance by leveraging complementary inductive biases. ViT emphasizes global token interactions, DeiT benefits from knowledge distillation, and Swin

captures local hierarchical structures via shifted windows.

Among all tested configurations, the *Random Forest stacking* attained the highest test accuracy of **85%**, surpassing the best single model (DeiT-Base-Distilled at 83.53%) by +2.08 percentage points. This result highlights the effectiveness of stacking for syndrome classification on this dataset.

Approach	Architecture/Meta	Learning bias	Acc. (%)
ViT-B/16	Transformer	Global token interactions	82.25
DeiT-Base-Distilled	Transformer (distilled)	Knowledge distillation from teacher	83.53
Swin-Tiny	Transformer (hierarchical)	Local shifted windows; hierarchical features	81.09
Stacking (Random Forest)	Meta-ensemble	Nonlinear aggregation of base logits	85
Stacking (Logistic Regression)	Meta-ensemble	Linear blending of model outputs	84.73
Stacking (XGBoost)	Meta-ensemble	Gradient-boosted trees over logits	83.84

Tableau 3.4: Comparison on the syndrome classification dataset (test accuracy, %).

Key Observations:

- **Complementary learning:** ViT captures global context, DeiT incorporates distilled supervisory signals, and Swin models local hierarchical features.
- **Stacking advantage:** The ensemble boosts performance by fusing diverse model representations. Random Forest stacking yields the best overall accuracy.
- **Both Random Forest and Logistic Regression stacking** outperform all single models, demonstrating robust cross-model complementarity.

Challenges: Despite the promising results, the study faced notable challenges due to the limited size and class imbalance of the dataset. The restricted number of syndrome-specific facial images constrained the model’s ability to generalize, increasing the risk of overfitting. Furthermore, variations in image quality, lighting conditions, and age distribution introduced additional noise, making the learning process more complex. These limitations underscore the importance of expanding and diversifying the dataset in future work to improve robustness and scalability of Transformer-based facial analysis systems.

3.6 conclusion

The proposed stacking framework—built on ViT, DeiT, and Swin transformer backbones and combined with classical meta-classifiers—provides a robust approach for syndrome classification from facial images. The modular design allows future extension (additional backbones, alternative meta-learners) and integration with interpretability methods to support clinical usage.

General Conclusion

Conclusion

In this thesis, the critical need for automated and accurate facial analysis in the early diagnosis of pediatric genetic disorders has been emphasized. Conventional diagnostic processes often rely heavily on clinical expertise and subjective visual assessments, which can lead to inconsistent results and delayed identification. To address this limitation, the proposed work introduced a transformer-based ensemble learning framework for facial syndrome classification, designed to improve diagnostic support systems through robust and intelligent automation.

The developed system integrated three advanced vision transformer architectures—Vision Transformer (ViT), Data-efficient Image Transformer (DeiT), and Swin Transformer—whose outputs were fused using an ensemble strategy based on meta-classifiers such as Random Forest, XGBoost, and Logistic Regression. Using a custom dataset consisting of facial images from both healthy individuals and patients with Down syndrome, the proposed model demonstrated strong performance and improved generalization capability. Among the meta-classifiers, the Random Forest achieved the best overall accuracy of **86%**, highlighting the effectiveness of the ensemble-based approach in capturing subtle phenotypic variations.

These findings confirm that transformer architectures, when combined through ensemble learning, can significantly enhance the performance of facial phenotype analysis systems, paving the way for practical clinical tools in genetic disorder screening.

Perspectives

Future research should explore several directions to build upon the results of this work:

- **Dataset expansion:** Extend the dataset to include a larger number of genetic syndromes and a more diverse population to improve model robustness and gener-

alization.

- **Multimodal integration:** Incorporate complementary data sources, such as genomic information or medical imaging, alongside facial features to enable a more comprehensive diagnostic framework.
- **Model interpretability:** Enhance explainability of the transformer-based models to better understand the decision-making process and increase trust among clinicians.
- **Clinical deployment:** Adapt and evaluate the system in real-world medical environments to assess its reliability and potential integration into healthcare workflows.

Appendix A

Annexe

A.1 Work Environment and Development Tools

A.1.1 Introduction

This section presents the hardware configuration, software environment, and libraries used for the development and experimentation of the proposed transformer-based ensemble learning system for pediatric syndrome diagnosis via facial analysis. The selected tools were chosen to ensure efficiency, scalability, and reproducibility throughout the project pipeline, from data preprocessing to model evaluation.

A.1.2 Hardware Tools

The entire implementation and experimentation were carried out using **Google Colab Pro**, a cloud-based platform providing GPU acceleration for deep learning tasks. The configuration used during training and evaluation is summarized below:

- **Processor:** Intel Xeon CPU @ 2.20GHz
- **GPU:** NVIDIA Tesla T4 (16 GB VRAM)
- **RAM:** 12 GB
- **Storage:** 100 GB Google Drive storage for dataset and checkpoint management

This setup allowed efficient training of multiple transformer architectures (ViT, DeiT, Swin) and ensemble learning models within reasonable computational time, leveraging GPU acceleration for faster convergence.

A.1.3 Software Tools

The development environment was based on the **Python** programming language, due to its extensive ecosystem for machine learning and deep learning. The following tools and frameworks were used:

- **Google Colab:** Used as the main development platform with GPU runtime for executing deep learning experiments.
- **PyTorch:** Main framework for building, training, and evaluating Vision Transformer (ViT), DeiT, and Swin Transformer architectures.
- **Hugging Face Transformers:** Provided pre-trained transformer backbones and model utilities for fine-tuning image classification tasks.
- **Split-Folders:** For automated dataset splitting into training, validation, and testing sets.
- **Scikit-learn:** For implementing ensemble learning algorithms (Random Forest, XGBoost, Logistic Regression) and performance evaluation.
- **XGBoost:** For gradient-boosted tree ensemble learning and comparison with Random Forest and Logistic Regression.
- **Google Drive:** Used for storing datasets, model checkpoints, and experiment outputs.
- **Matplotlib & Seaborn:** For plotting training curves, performance metrics, and confusion matrices.

A.1.4 Used Libraries

The following Python libraries and modules were integrated into the implementation pipeline:

- **torch, torchvision, torch.nn, torch.optim** — Deep learning model construction, optimization, and GPU acceleration.
- **transformers** — Access to pre-trained ViT, DeiT, and Swin Transformer models via Hugging Face.
- **timm** — Additional access to image transformer architectures and pretrained weights.

- **splitfolders** — For stratified splitting of the dataset into training, validation, and test sets.
- **numpy & pandas** — Data handling, numerical computation, and statistical operations.
- **matplotlib & seaborn** — Visualization of accuracy, loss trends, and confusion matrices.
- **xgboost** — Implementation of XGBoost-based meta-classifier for ensemble learning.
- **sklearn** — For Logistic Regression, Random Forest, data preprocessing, and evaluation metrics.
- **tqdm** — Progress visualization during training and evaluation loops.
- **zipfile & os** — File and dataset extraction, management, and automation.

A.1.5 Conclusion

In summary, the combination of cloud-based GPU resources, Python’s flexible development environment, and specialized deep learning libraries provided an optimal setup for this project. The integration of PyTorch and Hugging Face enabled seamless implementation of transformer-based architectures, while Scikit-learn and XGBoost facilitated robust ensemble learning and evaluation. This cohesive environment ensured efficient experimentation, reproducibility, and reliable performance of the proposed framework.

A.2 Implementation

This section presents the detailed implementation process of the proposed Transformer-based ensemble framework for pediatric genetic syndrome diagnosis. The pipeline consists of thirteen major steps, from data preparation to final evaluation.

Step 1: Importing Required Libraries

This step imports the necessary Python libraries for data handling, model building, and evaluation.

```
1 import os
2 import numpy as np
3 import pandas as pd
4 import tensorflow as tf
5 from tensorflow.keras.preprocessing.image import ImageDataGenerator
6 from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
7 from sklearn.linear_model import LogisticRegression
8 from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

Listing A.1: Importing Libraries

Step 2: Loading and Organizing the Dataset

Images are loaded from the dataset directory and split into training, validation, and test sets.

```
1 dataset_dir = "/content/dataset"
2
3 train_datagen = ImageDataGenerator(
4     rescale=1./255,
5     validation_split=0.30
6 )
7
8 train_gen = train_datagen.flow_from_directory(
9     dataset_dir,
10    target_size=(96, 96),
11    batch_size=32,
12    subset='training',
13 )
14
15 val_gen = train_datagen.flow_from_directory(
16    dataset_dir,
17    target_size=(96, 96),
```

```
18     batch_size=32,  
19     subset='validation'  
20 )
```

Listing A.2: Dataset Loading and Splitting

Step 3: Data Preprocessing

Image augmentation enhances diversity and helps avoid overfitting.

```
1 train_aug = ImageDataGenerator(  
2     rescale=1./255,  
3     rotation_range=15,  
4     zoom_range=0.1,  
5     horizontal_flip=True  
6 )
```

Listing A.3: Data Preprocessing

Step 4: Building the Vision Transformer (ViT) Model

The Vision Transformer (ViT) model is initialized and fine-tuned using pre-trained weights.

```
1 vit_model = tf.keras.applications.vit.ViT(  
2     input_shape=(96,96,3),  
3     include_top=False,  
4     weights='imagenet'  
5 )
```

Listing A.4: Vision Transformer Model

Step 5: Building the DeiT Model

The Data-efficient Image Transformer (DeiT) adapts Transformer performance to smaller datasets.

```
1 deit_model = tf.keras.applications.deit.DeiT(  
2     input_shape=(96,96,3),  
3     include_top=False,  
4     weights='imagenet'  
5 )
```

Listing A.5: DeiT Model

Step 6: Building the Swin Transformer Model

The Swin Transformer (Swin-Tiny) captures hierarchical visual representations.

```
1 swin_model = tf.keras.applications.swin.SwinTransformer(  
2     input_shape=(96,96,3),  
3     include_top=False,  
4     weights='imagenet'  
5 )
```

Listing A.6: Swin Transformer Model

Step 7: Model Compilation

Each model is compiled with the Adam optimizer and categorical cross-entropy loss.

```
1 for model in [vit_model, deit_model, swin_model]:  
2     model.compile(optimizer='adam',  
3                 loss='categorical_crossentropy',  
4                 metrics=['accuracy'])
```

Listing A.7: Model Compilation

Step 8: Model Training

Each model is trained individually on the prepared dataset.

```
1 history_vit = vit_model.fit(train_gen, validation_data=val_gen, epochs=30)  
2 history_deit = deit_model.fit(train_gen, validation_data=val_gen, epochs=30)  
3 history_swin = swin_model.fit(train_gen, validation_data=val_gen, epochs=30)
```

Listing A.8: Model Training

Step 9: Feature Extraction

The output probabilities from the models are extracted for ensemble training.

```
1 vit_preds = vit_model.predict(val_gen)  
2 deit_preds = deit_model.predict(val_gen)  
3 swin_preds = swin_model.predict(val_gen)  
4  
5 X_meta = np.concatenate([vit_preds, deit_preds, swin_preds], axis=1)  
6 y_meta = val_gen.classes
```

Listing A.9: Feature Extraction

Step 10: Ensemble Model Construction

Meta-classifiers (Random Forest, XGBoost, and Logistic Regression) are used for fusion.

```
1 rf = RandomForestClassifier(n_estimators=200)
2 xgb = GradientBoostingClassifier()
3 lr = LogisticRegression(max_iter=500)
```

Listing A.10: Meta Classifier Construction

Step 11: Training the Meta-Classifier

The meta-classifiers are trained on the combined features.

```
1 rf.fit(X_meta, y_meta)
2 xgb.fit(X_meta, y_meta)
3 lr.fit(X_meta, y_meta)
```

Listing A.11: Training Meta Classifiers

Step 12: Evaluation on Test Set

The models are tested on unseen data to measure performance.

```
1 test_preds = np.concatenate([
2     vit_model.predict(test_gen),
3     deit_model.predict(test_gen),
4     swin_model.predict(test_gen)
5 ], axis=1)
6
7 rf_acc = accuracy_score(test_gen.classes, rf.predict(test_preds))
8 print("Random Forest Accuracy:", rf_acc)
```

Listing A.12: Evaluation on Test Data

Step 13: Results and Visualization

Results such as confusion matrix and classification report are displayed.

```
1 print("Classification Report:\n", classification_report(test_gen.classes, rf.predict(
2     test_preds)))
3
4 cm = confusion_matrix(test_gen.classes, rf.predict(test_preds))
5 plt.imshow(cm, cmap='Blues')
6 plt.title("Confusion Matrix")
```

```
6 plt.xlabel("Predicted")
7 plt.ylabel("True")
8 plt.show()
```

Listing A.13: Results Visualization

Bibliography

- Belhacini, A. (2025). *User’s experimental implementation using mobilenetv2 on a pediatric facial dataset*. (Unpublished experimental study)
- Bojarski, M., et al. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*(3), 235–255.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
- Bull, M. J. (2020). Down syndrome. *New England Journal of Medicine*, *382*(24), 2344–2352. doi: 10.1056/NEJMra1500527
- Böhmer, K. D., et al. (2020). Impact of image quality on automated facial recognition in genetic syndromes. *European Journal of Human Genetics*, *28*(4), 590–598.
- Cardaioli, S., et al. (2023). Cnn–transformer hybrids for interpretable syndrome classification from facial images. *Artificial Intelligence in Medicine*, *138*, 102405.
- Chambino, P., et al. (2021). Facial phenotyping with generative adversarial networks for rare syndrome detection. In *Proceedings of the IEEE international conference on computer vision (iccv) workshops*.
- Cohen, M. M. (2017). Apert syndrome update, differential diagnosis, and epidemiology. *American Journal of Medical Genetics Part A*, *173*(6), 1627–1637. doi: 10.1002/ajmg.a.38112
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2020). An image is worth 1616 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

- Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. Wiley.
- Esteva, A., Kuprel, B., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.
- Ge, Y., et al. (2022). Dual-branch attention networks for facial syndrome classification. *IEEE Transactions on Medical Imaging*, *41*(2), 301–312.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Guo, T., Dong, J., Li, H., & Gao, Y. (2017). Simple convolutional neural network on image classification. In *Proceedings of the 2017 IEEE 2nd international conference on big data analysis (icbda)* (pp. 721–724). IEEE. doi: 10.1109/ICBDA.2017.8078730
- Gurovich, Y., Hanani, Y., Bar, O., et al. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, *25*(1), 60–64. doi: 10.1038/s41591-018-0279-0
- Hangaragi, K., et al. (2023). Attention-based cnn models for genetic syndrome classification from facial features. *Journal of Medical Genetics*, *60*(7), 589–596.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 770–778).
- Hinton, G., Deng, L., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, *29*(6), 82–97.
- Hsieh, T. H., et al. (2022). Gestaltmatcher: Overcoming the limits of rare disease matching using facial phenotypes. *American Journal of Human Genetics*, *109*(5), 876–888.
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 7132–7141).
- Hunter, A., Smith, R., & Jones, P. (2018). Challenges in automated recognition of facial asymmetry in goldenhar syndrome. *Journal of Medical Imaging*, *5*(3), 1–9.
- Ikromovich, N., et al. (2023). Real-time syndrome diagnosis using lightweight deep learning models on mobile devices. *Computers in Medicine and Biology*, *146*, 106829.
- Javatpoint. (n.d.). *Machine learning: Support vector machine algorithm*. <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.
- Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 1867–1874).
- Kline, A. D., Krantz, I. D., et al. (2020). Cornelia de lange syndrome: Clinical review,

- diagnostic criteria, and research directions. *American Journal of Medical Genetics Part C*, 142(3), 158–166. doi: 10.1002/ajmg.c.30202
- Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, 32(11), 1238–1274.
- Kumar, A., et al. (2023). Guided efficientnet for facial analysis in genetic syndrome prediction. *Computers in Biology and Medicine*, 158, 106764.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Liu, X., et al. (2023). Multi-label classification of facial dysmorphology for genetic disorders using resnet. *Medical Image Analysis*, 82, 102620.
- Lumaka, A., et al. (2017). Facial dysmorphism is influenced by ethnic background of the patient and evaluator. *Clinical Genetics*, 92(1), 166–171.
- McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *Aaai workshop on learning for text categorization* (Vol. 752, pp. 41–48).
- Mikolov, T., Sutskever, I., et al. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems (neurips)* (Vol. 26).
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4), 875–889.
- Pacheco, J. (2021a). *Introduction to machine learning*. (CSC 380 Course Materials)
- Pacheco, J. (2021b). *Introduction to machine learning [lecture notes for csc380, fall 2021]*. http://www.pacheco.j.com/courses/csc380_fall21/lectures/mlintro.pdf.
- Quinlan, J. R. (1996). Improved use of continuous attributes in c4.5. *Journal of Artificial Intelligence Research*, 4, 77–90.
- Rahman, M., et al. (2023). Augmenting rare syndrome datasets with synthetic faces for improved deep learning diagnosis. *Expert Systems with Applications*, 207, 118062.
- Ricci, F., Rokach, L., et al. (2011). *Recommender systems handbook*. Springer.
- Riedmiller, M. (2005). Neural fitted q iteration—first experiences with a data efficient neural reinforcement learning method. In *European conference on machine learning (ecml)* (pp. 317–328).
- Rodriguez, D., et al. (2024). Federated learning for collaborative facial syndrome diagnosis across hospitals. In *Proceedings of the ieee engineering in medicine and biology conference (embc)*.
- Russell, S., & Norvig, P. (2020a). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

- Russell, S., & Norvig, P. (2020b). *Artificial intelligence: A modern approach* (4th ed.). Hoboken, NJ, USA: Pearson.
- Sallab, A. E. S., Abdou, M., et al. (2017). Deep reinforcement learning framework for autonomous driving. *Electronic Imaging, 2017*(19), 70–76.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (cvpr)* (pp. 4510–4520).
- Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis*. Wiley.
- Selvaraju, R. R., et al. (2020). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision, 128*(2), 336–359.
- Shah, S. N., & McKenna, P. J. (2021). Machine learning in pediatric genetic diagnosis: Progress and challenges. *Frontiers in Genetics, 12*, 676123. doi: 10.3389/fgene.2021.676123
- Silver, D., Schrittwieser, J., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature, 529*(7587), 484–489.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th international conference on machine learning (icml)*.
- Tomašević, N., et al. (2024). Noise-resilient deep learning models for robust facial diagnosis. *IEEE Access, 12*, 3498–3511.
- TpointTech. (n.d.). *K-nearest neighbor algorithm for machine learning*. <https://www.tpointtech.com/k-nearest-neighbor-algorithm-for-machine-learning>.
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE computer vision and pattern recognition conference (cvpr)* (Vol. 1, pp. 511–518).
- Williams, J. D., Asadi, K., & Zweig, G. (2017). Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning. In *Proceedings of the 55th annual meeting of the association for computational linguistics (acl)*.
- Williams, S., Johnson, E., & White, S. (2019). Age-related phenotypic variation in angel-

- man syndrome: Implications for diagnosis and intervention. *Genetics in Medicine*, 21(5), 1041–1050. doi: 10.1038/s41436-018-0407-1
- Yadav, D. (n.d.). *Different machine learning models*. <https://medium.com/@yadavdeepika729/different-machine-learning-models-2387575e64cd>. (Medium)
- Yu, C. N., et al. (2019). Reinforcement learning for healthcare: Challenges, opportunities, and future directions. In *Neurips workshop on machine learning for health (ml4h)*.
- Zhang, X., et al. (2022). A multi-syndrome facial classification system using deep learning. *Journal of Medical Genetics*, 59(1), 45–52.
- Zhang, Z., Zhu, Y., Liu, M., Zhang, Z., Zhao, Y., Yang, X., . . . Zhang, L. (2022). Artificial intelligence-enhanced echocardiography for systolic function assessment. *Journal of Clinical Medicine*, 11(10), 2893. doi: 10.3390/jcm11102893



جامعة 20 أوت سكيكدة
كلية العلوم
قسم الإعلام الآلي



مشروع لنيل شهادة ماستر + شهادة مؤسسة ناشئة في اطار القرار الوزاري 1275

نظام ذكي للتشخيص المبكر للمتلازمات الوراثية لدى الأطفال باستخدام تحليل ملامح الوجه

من إعداد الطلبة:

بلحاسيني أميلة
لموشي آلاء رحمة

التخصص : ذكاء إصطناعي

لجنة المناقشة:

رئيس اللجنة: د بن ديب رياض

الأستاذ المشرف: د.حزمون سميرة

الأستاذ الممتحن: د حمروش هدى

خبير BMC:

الشريك الإقتصادي: لمين سوداني

السنة الجامعية: 2024 / 2025

الفهرس

V	قائمة الصور
VI	قائمة الجداول
VII	إهداء
VIII	شكر و عرفان
IX	مقدمة
1	1 تقديم المشروع
1	1.1 تمهيد
1	2.1 فكرة المشروع
1	1.2.1 كيف بدأت الفكرة
2	2.2.1 ما الذي سوف نقوم به
2	3.2.1 القيم المقترحة
4	3.1 فريق العمل
5	4.1 أهداف المشروع
6	5.1 الجدول الزمني
6	1.5.1 التخطيط والتحضير
6	2.5.1 التصميم و التطوير
6	3.5.1 الإطلاق والتسويق
6	4.5.1 الدعم والتحسين المستمر
8	2 الجوانب الإبتكارية
8	1.2 طبيعة الإبتكارات

8	مجالات الابتكار	2.2
8	تطبيق تكنولوجيا متقدمة	1.2.2
9	نموذج الأعمال المبتكر	2.2.2
9	العروض والخصومات	3.2.2
10	التحليل البياني	4.2.2
10	التكامل الشامل	5.2.2
10	الاستدامة البيئية	6.2.2
12	3 التحليل الاستراتيجي للسوق	
12	عرض القطاع السوقي	1.3
12	السوق المحتمل	1.1.3
12	أعداد السوق المحتملين	2.1.3
13	السوق المستهدف	3.1.3
13	التحليل الاستراتيجي للسوق	2.3
13	من سيستخدم جهازنا؟	1.2.3
14	من سيستخدم تطبيقنا؟	2.2.3
14	دوافع الاستخدام	3.2.3
14	مبررات اختيار هذا السوق	4.2.3
15	إمكانية إبرام عمليات شراء	5.2.3
15	قياس شدة المنافسة في السوق الجزائرية	3.3
15	المنافسون المباشرون	1.3.3
15	المنافسون غير المباشرين	2.3.3
16	التحليل العام للمنافسة	3.3.3
16	الإستراتيجية التسويقية	4.3
16	تحليل السوق والفهم العميق للزبائن	1.4.3
16	تحديد الميزة التنافسية	2.4.3
17	استخدام قنوات التسويق الفعالة	3.4.3
17	تحسين تجربة المستخدم	4.4.3
17	استراتيجيات الاحتفاظ بالعملاء	5.4.3
18	الشراكات الاستراتيجية	6.4.3

18 7.4.3 تقييم الأداء والتحسين المستمر

4 خطة الإنتاج والتنظيم 19

19 1.4 خطة الإنتاج

19 1.1.4 التخطيط والتحضير

19 2.1.4 التصميم والتطوير

20 3.1.4 الإطلاق والتسويق

20 4.1.4 الدعم والتحسين المستمر

20 5.1.4 التقييم والاستدامة

21 2.4 الجدول الزمني

5 الخطة التسويقية 22

22 1.5 الخطة المالية للمشروع ونموذج العمل التجاري

22 1.1.5 التكاليف والرسوم

24 2.5 تمويل المشروع

25 3.5 رقم الأعمال

25 1.3.5 النظرة التفاؤلية

25 2.3.5 النظرة التشاؤمية

26 4.5 جداول حسابات النتائج

27 5.5 خطة الخزينة

6 النموذج التجريبي الأولي 29

30 1.6 التطبيق

30 1.1.6 صفحة البدء

32 2.1.6 صفحة اختيار اللغة

33 3.1.6 صفحة الترحيب

34 4.1.6 صفحة التسجيل

35 5.1.6 إنشاء الحساب

36 6.1.6 الصفحة الرئيسية

37 7.1.6 صفحة التعليمات

38 8.1.6 صفحة من نحن

40	البنية التقنية لتطبيق NeoFaceMed	9.1.6
41	نموذج الجهاز التطبيقي	2.6
41	مخطط توصيل المكونات الإلكترونية	1.2.6
42	الواجهة الأمامية للجهاز	2.2.6
42	الواجهة الخلفية للجهاز	3.2.6
44	نموذج العمل التجاري	7
45	خاتمة	

قائمة الصور

30	1.6 واجهة البدء
32	2.6 صفحة اختيار اللغة
33	3.6 صفحة الترحيب
34	4.6 واجهة إنشاء الحساب
35	5.6 واجهة إنشاء الحساب
36	6.6 الصفحة الرئيسية
37	7.6 صفحة التعليمات
38	8.6 صفحة من نحن
39	9.6 اكتشف المزيد
41	10.6 مخطط توصيل المكونات الإلكترونية
42	11.6 الجهاز من الأمام
42	12.6 الجهاز من الخلف
44	1.7 نموذج العمل التجاري

قائمة الجداول

7	الجدول الزمني لإطلاق التطبيق	1.1
21	الجدول الزمني التفصيلي لإطلاق التطبيق	1.4
24	جدول التكاليف والإستثمار (التوقعات)	1.5
25	تطور الإيرادات المتوقعة من بيع الجهاز الذكي (النظرة التفاؤلية)	2.5
25	تطور الإيرادات المتوقعة من التطبيق الهاتفي (النظرة التفاؤلية)	3.5
26	تطور المبيعات والإيرادات للجهاز – النظرة التشاؤمية	4.5
26	تطور الاشتراكات والإيرادات للتطبيق – النظرة التشاؤمية	5.5
26	جدول المصاريف (الأشهر من 1 إلى 6)	6.5
27	جدول المصاريف (الأشهر من 7 إلى 12)	7.5
27	الإيرادات والنفقات المتوقعة خلال السنة الأولى (الأشهر 1 إلى 6)	8.5
27	الإيرادات والنفقات المتوقعة خلال السنة الأولى (الأشهر 7 إلى 12)	9.5
28	الإيرادات والنفقات المتوقعة خلال السنة الخامسة (الأشهر 1 إلى 6)	10.5
28	الإيرادات والنفقات المتوقعة خلال السنة الخامسة (الأشهر 7 إلى 12)	11.5

إهداء

نهدي ثمرة جهدنا المتواضع إلى من كانوا بعد الله السند والعون في مسيرتنا،
إلى والدينا العزيزين، نبض القلب ونور الدرب، جزاهما الله عنا خير الجزاء وأطال في عمرهما على الطاعة والصحة
والعافية.

إلى إخوتنا وأخواتنا، الذين كانوا دومًا مصدر التشجيع والدعم والمساندة.
إلى أهلنا وأقاربنا الكرام، على دعواتهم الصادقة ومساندتهم المستمرة.
إلى أساتذتنا الأفاضل، من غرسوا فينا حب العلم، وكانوا لنا قدوة في العطاء والاجتهاد.
إلى زملائنا وزميلاتنا، الذين جمعنا بهم سنوات من السعي والمثابرة والمواقف التي لا تُنسى.
إلى كل من قدّم لنا يد العون أو كلمة دعم أو ابتسامة أمل في طريق هذا العمل.
وإلى كل باحثٍ عن المعرفة، يؤمن أن بالعلم تُبنى الأمم وتُصان الكرامة.

شكر و عرفان

قال تعالى: رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وَالِدَيَّ وَأَنْ أَعْمَلَ صَالِحًا تَرْضَاهُ وَأَدْخِلْنِي
بِرَحْمَتِكَ فِي عِبَادِكَ الصَّالِحِينَ (19) سورة النمل

الحمد لله الذي بنعمته تتم الصالحات، وبفضله تنيسر الخطوات، وبعونه اكتمل هذا الجهد المتواضع، فله الحمد حمداً يليق
بجلال وجهه وعظيم سلطانه.

نتقدّم بأسمى آيات الشكر والعرفان إلى أستاذتنا الفاضلة ”حزمون سميرة“، لما أولته من اهتمام ومتابعة دقيقة، وتوجيهات
قيّمة، ودعم متواصل كان له الأثر البالغ في أن يرى هذا المشروع النور بأبهى صورة. فجزاها الله عنا خير الجزاء، ووفقها
لما فيه مزيد من التآلق والعطاء.

كما نتوجّه بخالص الامتنان إلى والدينا العزيزين، اللذين كانا لنا السند والعون، بدعواتهما الصادقة، وتضحياتهما الجليلة،
فلهما منا كل الحب والتقدير والعرفان ما حيننا.

ولا ننسى أن نخصّ بالشكر عائلتنا الكريمة التي أحاطتنا بالدعم المعنوي، والتشجيع الدائم، فكانت مصدر طاقة وإلهام خلال
مسيرتنا.

ونتوجّه أيضاً بالشكر إلى جميع الأساتذة الذين أسهموا في تكويننا العلمي، وإلى زملائنا الذين شاركونا رحلة الدراسة بما فيها
من جهد وتعاون، وإلى كل من قدّم لنا يد العون أو كلمة دعم صادقة خلال هذه التجربة الأكاديمية.

فلكم جميعاً منا أصدق الدعاء وأسمى عبارات الشكر والامتنان.

مقدمة

في عصر يشهد تطوراً غير مسبوق في مجال الذكاء الاصطناعي والتقنيات الرقمية، بات من الضروري توظيف هذه الأدوات الحديثة لخدمة القضايا الإنسانية والاجتماعية، وعلى رأسها الرعاية الصحية للأطفال. يمثل التشخيص المبكر للأمراض المزمنة والجينية لدى الأطفال أحد أهم التحديات التي تواجه الأنظمة الصحية في الدول النامية والمتقدمة على حد سواء، إذ أن التأخر في اكتشاف هذه الأمراض قد يؤدي إلى مضاعفات جسدية ونفسية يصعب تداركها لاحقاً.

يعاني آلاف الأطفال سنوياً من أمراض مثل متلازمة داون، فقر الدم المنجلي، متلازمات وراثية نادرة، وحتى الاضطرابات العصبية والسلوكية مثل طيف التوحد، دون أن يتم تشخيصهم في الوقت المناسب. هذا التأخير لا يعود فقط إلى ضعف الإمكانيات الطبية، بل أيضاً إلى غياب أدوات الفحص المبدي الذكية التي يمكن أن تُستخدم في الخطوط الأمامية للرعاية، كالمراكز الصحية ومراكز التلقيح.

من هذا المنطلق جاءت فكرة مشروع **NeoFaceMed** جهاز وتطبيق هاتفي يعتمدان على تحليل ملامح الوجه باستخدام تقنيات رؤية الحاسوب والتعلم العميق للكشف المبكر عن مجموعة من الأمراض الوراثية والمزمنة التي تترك بصمتها على شكل الوجه وبنية وتناسقه وتعبيراته. هذه التقنية ليست بديلاً عن التشخيص الطبي المتخصص، لكنها أداة مساعدة قوية تُنبئ الفريق الطبي إلى وجود مؤشرات تستوجب الفحص الدقيق. ويعدّ هذا الفحص الذكي من الأول من نوعه الذي سيُدمج في دور الأمومة والطفولة بالتوازي مع حملات التلقيح الدورية، بما يضمن وصول هذه التقنية إلى أكبر عدد ممكن من الأطفال في مرحلة مبكرة من حياتهم، ويتيح فرصة ثمينة لاكتشاف الحالات قبل ظهور الأعراض الواضحة. إن أهمية المشروع لا تكمن في جانبه التقني فحسب، بل تتجلى أيضاً في بُعد الإنسان والاجتماعي؛ فهو مشروع يدعم الأمهات والآباء ويمنحهم الطمأنينة، ويعزز العدالة الصحية عبر توفير وسيلة فحص أولية موحدة لجميع الأطفال دون تمييز، ويسهم في تخفيف العبء عن الطاقم الطبي وتقليل حالات التشخيص المتأخر المُكلفة على المدى البعيد.

بذلك يمثل **NeoFaceMed** نقلة نوعية في وصل الذكاء الاصطناعي بصحة الأطفال، ويضع أسساً عملية لتعميم الفحص الذكي المبكر في نقاط الرعاية الأساسية، ويفتح آفاقاً مستقبلية لتحسين جودة الحياة وحماية الأجيال القادمة من المعاناة الصامتة.

المحور 1

تقديم المشروع

1.1 تمهيد

يمثل التشخيص المبكر للأمراض الوراثية التي تظهر على ملامح الوجه خطوة حاسمة في توفير الرعاية الصحية المناسبة في الوقت المناسب.

مع تقدم الذكاء الاصطناعي وتوسع استخدامه في القطاع الصحي، أصبح من الممكن اليوم تطوير أدوات تقنية تساعد على اكتشاف بعض المؤشرات الوراثية من خلال ملامح الوجه، وهي خطوة يمكن أن تُحدث فرقًا كبيرًا في حياة الأطفال والأسر والمجتمع ككل.

مشروعنا يقترح تطوير نظام ذكي يقوم بتحليل ملامح وجه الطفل بهدف رصد مؤشرات قد تدل على وجود أمراض وراثية، ويهدف إلى دمج كفحص إلزامي في مراحل التلقيح المبكرة بالمراكز الصحية لضمان فحص جميع الأطفال بطريقة شاملة ومبكرة، مما يسهل التدخل السريع والفعال عند الحاجة.

2.1 فكرة المشروع

1.2.1 كيف بدأت الفكرة

بدأت فكرة هذا المشروع من ملاحظة علمية وطبية مؤكدة، وهي أن العديد من الأمراض الوراثية تترك آثارًا ملموسة على ملامح وجه الإنسان، خصوصًا في مرحلة الطفولة المبكرة. هذه العلامات تشمل على سبيل المثال لا الحصر:

- تموضع غير معتاد للعينين أو شكل الجبهة،
- تفاوت في التناسق الوجهي،
- سمات معينة في الأنف أو الفم،

• تغيّرات تعبيرية مزمنة مرتبطة بأمراض عصبية أو مناعية أو دموية مزمنة.

ومع أن هذه السمات قد تكون واضحة في بعض الحالات، فإن اكتشافها يتطلب غالبًا تدريبًا طبيًا متخصصًا، مثل اختصاصيي الوراثة أو أطباء الأطفال ذوي الخبرة، وهو ما لا يتوفر دائمًا، خصوصًا في المناطق النائية أو ذات الموارد الطبية المحدودة. في الواقع، فإن جزءًا كبيرًا من الأطفال المصابين بمتلازمات وراثية أو أمراض مزمنة لا يتم تشخيصهم إلا في مراحل متأخرة، بعد أن تبدأ الأعراض في التفاقم أو تظهر مضاعفات واضحة، ما يؤدي إلى ضياع ”النافذة الذهبية” للتدخل المبكر. من هنا ظهرت الحاجة إلى حلّ ذكي يمكن استخدامه في الخطوط الأولى للرعاية الصحية، أي في المراكز الصحية العامة، وعيادات التفقيح، والوحدات الجوارية، حيث يتعامل الطاقم الطبي مع مئات الأطفال يوميًا، دون أن تكون لديهم الوسائل أو الوقت الكافي لفحص كل حالة بدقة تخصصية.

وجاءت الفكرة: لماذا لا نعتمد على الذكاء الاصطناعي، وخوارزميات التعلم العميق، لتحليل صور وجوه الأطفال والتقاط المؤشرات البصرية الدقيقة التي قد تدل على هذه الأمراض؟ إذا كان الإنسان المتخصص قادرًا على تمييز هذه العلامات بالعين المجردة، فإن الحاسوب يمكن تدريبه على القيام بالأمر نفسه – بل بسرعة ودقة واستمرارية – دون تأثر بالتعب أو ضغط العمل.

2.2.1 ما الذي سوف نقوم به

سيتم تطوير جهاز ذكي محمول مخصّص للاستعمال داخل المراكز الصحية ومراكز التفقيح، مدمج بكاميرا بسيطة وشاشة صغيرة، بالإضافة إلى تطبيق هاتفي مكمل موجه للأولياء والأطباء. يقوم كلا النظامين بالتقاط صورة لوجه الطفل ثم معالجتها مباشرة باستخدام خوارزميات الذكاء الاصطناعي وتقنيات رؤية الحاسوب. سيساعد النظام على اكتشاف مؤشرات أولية لاحتمال الإصابة بعدد واسع من المتلازمات والأمراض الجينية المزمنة التي تترك بصمتها على ملامح الوجه، ومن بينها:

• متلازمة تيرنر (Turner Syndrome)،

• متلازمة ويليامز (Williams Syndrome)

• متلازمة أنجلمن (Angelman Syndrome)

• متلازمة نونان (Noonan Syndrome)

وظيفة النظام ليست التشخيص النهائي، بل تقديم تنبيه للطبيب أو الممرضة بوجود ”احتمال مرتفع” يدعو إلى التحقق أو تحويل الطفل إلى مختص. كما سيتضمن التطبيق واجهة مبسطة، ومزايا لحفظ وتحليل البيانات بشكل آمن، مع مراعاة خصوصية المستخدمين.

3.2.1 القيم المقترحة

مشروع ”NeoFaceMed” لا يقَدّم فقط حلاً تقنيًا، بل يحمل في جوهره رؤية إنسانية شاملة تقوم على مجموعة من القيم الجوهرية التي تشكل أساس تصميمه وتنفيذه وتطويره المستمر. هذه القيم ليست شعارات، بل مبادئ عملية موجّهة لجميع

القرارات التقنية والطبية والتنظيمية في كل مرحلة من المشروع.

• العدالة الصحية (Equity in Health)

يهدف المشروع إلى كسر الحواجز الاجتماعية والجغرافية من خلال تقديم أداة ذكية وموحدة للفحص المبكر، تضمن أن يخضع كل طفل للفحص الإلزامي بغض النظر عن مكان ولادته أو مستوى دخل أسرته. هذا يضمن تكافؤ فرص الاكتشاف المبكر والرعاية اللاحقة، فالعدالة الصحية ليست ترفاً، بل واجب أخلاقي وركيزة من ركائز أهداف التنمية المستدامة في مجال الصحة.

• الوقاية قبل العلاج (Prevention First)

يؤمن المشروع بأن التدخل المبكر يوفر حياة أكثر جودة وأقل كلفة، سواء من حيث المعاناة الفردية أو العبء المالي على الأسرة والنظام الصحي. بدل انتظار ظهور الأعراض المتقدمة، يقدم النظام أداة استباقية لاكتشاف المؤشرات الأولى للمرض، مما يسمح باتخاذ قرارات علاجية مبكرة وفعالة.

• تمكين الأسرة (Empowering Families)

الأمهات والآباء هم الشركاء الأوائل في رعاية الطفل، لكنهم غالباً ما يجدون أنفسهم في حالة قلق أو غموض تجاه حالة طفلهم. يوفر التطبيق وسيلة بسيطة تمنحهم وضوحاً مبدئياً، وتعزز ثقتهم في النظام الصحي، وتشركهم في المتابعة. كل تنبيه أو توصية يقدمها النظام هو نقطة انطلاق لحوار مسؤول بين الأسرة والطبيب.

• تعزيز الكفاءة الطبية (Medical Support Optimization)

الطاقم الطبي، خصوصاً في المناطق ذات الضغط العالي، يواجه تحديات كبيرة في فحص مئات الأطفال يومياً. يهدف المشروع إلى دعم الطبيب بأداة دقيقة وسريعة تساعده على اتخاذ قرارات مبنية على بيانات وتحليل علمي، خاصة في الحالات التي لا تظهر أعراضها بوضوح.

• الخصوصية والكرامة (Privacy and Dignity)

يولي المشروع أهمية قصوى لحماية خصوصية الأطفال وكرامتهم، إذ يرتبط الأمن القومي ارتباطاً وثيقاً بأمن المعلومات وسلامة البيانات الصحية. يلتزم بتطبيق أعلى معايير الأمن والتشفير لضمان سرية البيانات ومنع أي استخدام غير مشروع، مع إنشاء قاعدة بيانات وطنية للأطفال الجزائريين تُدار محلياً تحت إشراف السلطات الصحية المختصة، تعزيزاً للسيادة الرقمية وحماية للمعطيات وفق الضوابط القانونية والأخلاقية.

• السيادة الرقمية والمصلحة الوطنية (Digital Sovereignty & National Interest)

المشروع لا يهدف فقط إلى تقديم خدمة طبية ذكية، بل أيضاً إلى المساهمة في بناء استقلالية رقمية في مجال البيانات الحيوية داخل منطقة المغرب العربي. في ظل ما يُعرف اليوم بـ"حرب البيانات"، تصبح الدول التي تملك بياناتها وتتحكم في معالجتها أكثر قدرة على اتخاذ قرارات علمية وسياسية مستقلة. من خلال إدماج النظام في مؤسسات الدولة والمراكز البحثية الوطنية، يساهم المشروع في تقوية البنية التحتية للبيانات الطبية، ويجعل من الجزائر نموذجاً رائداً

في حماية مواردها الرقمية واستثمارها لصالح المجتمع والبحث العلمي المحلي. إن امتلاك هذه البيانات وتنظيمها بشكل مؤسسي سيمنح الدولة قوة معرفية واستراتيجية في تطوير سياساتها الصحية والتعليمية والبحثية المستقبلية.

• الابتكار الموجه نحو الإنسان (Human-Centered Innovation)

تم تصميم كل عنصر في النظام — من الواجهة إلى طريقة الإشعار والتحليل — بناءً على احتياجات المستخدمين الفعليين: الأطباء، المرضى، والوالدين. هذا النهج يُعزز قابلية الاستخدام وفعالية النتائج ويقلل من مقاومة التبني الميداني للتكنولوجيا الجديدة.

• الاستدامة والتطوير المستمر (Sustainable Impact)

القيمة لا تكمن فقط في إطلاق البرنامج، بل في قدرته على التكيف والنمو. المشروع مصمم بحيث يكون قابلاً للتحديث المستمر، بإضافة أمراض جديدة، وتحسين الخوارزميات، والتوسع جغرافياً، بالشراكة مع مؤسسات تعليمية وطبية، لضمان بقاء تأثيره طويل الأمد.

خاتمة: إن مشروع "NeoFaceMed" يجسد التقاء التقنية بالإنسانية والسيادة الوطنية؛ فهو ليس مجرد نظام ذكي للكشف المبكر، بل مبادرة تسعى لترسيخ قيم العدالة، والابتكار، والاستقلالية الرقمية، بما يخدم مستقبل الصحة في الجزائر والمغرب العربي بأكمله.

3.1 فريق العمل

يتكون فريق المشروع من الطالبتين بلحاسيني أميلة، لموشي آلاء رحمة. الفريق مسؤول عن تطوير خوارزميات الذكاء الاصطناعي والتحسين، إضافة إلى ذلك تم توزيع المهام الجانبية بينهم كما يلي:

1. المهام الجانبية للطالبة بلحاسيني أميلة:

- جمع البيانات اللازمة وإعدادها للتدريب والتعامل مع قواعد البيانات.
- تنظيف وتحليل البيانات لضمان جودتها وصحتها.
- تطوير وصيانة قاعدة البيانات المستخدمة لتخزين البيانات.
- تنفيذ عمليات تحويل البيانات والمعالجة المسبقة لجعلها جاهزة للتعلم الآلي.
- التسويق والإشهار للمشروع، والبحث عن الشركاء الاقتصاديين والمستثمرين.

2. المهام الجانبية للطالبة لموشي آلاء رحمة :

- تصميم واجهة المستخدم وتطوير التطبيق.
- تطوير النماذج التفاعلية وتجربة المستخدم.

- اختبار البرمجيات لضمان جودتها ووظائفها.
 - التعاون مع فريق التصميم لضمان توافق الواجهة مع متطلبات المستخدم.
 - إدارة تحديثات التطبيق وإصلاح الأخطاء.
 - التواصل مع الشركات والمستثمرين المحتملين لعرض المشروع وجذب الدعم المالي.
- بهذه الطريقة يتم توزيع المهام بشكل يضمن تكامل الجهود بين أعضاء الفريق لتحقيق أهداف المشروع بنجاح.

4.1 أهداف المشروع

- تطوير نظام ذكي لتحليل ملامح الوجه باستخدام الذكاء الاصطناعي للكشف المبكر عن الأمراض الوراثية.
- استخدام الفحص الوجهي كأداة أولية تُمكن من التعرف على مؤشرات المرض بسرعة ودون تدخل طبي متخصص مباشر.
- دمج النظام في مراحل التلقيح والرعاية الصحية الأولية لضمان فحص كل الأطفال بشكل تلقائي ومنهجي.
- التقليل من زمن التأخير في التشخيص وتحسين فرص التدخل المبكر والعلاج المناسب قبل تطور الحالة.
- تمكين الأسرة بمعلومة طبية أولية تسهل عليها اتخاذ قرارات مدروسة بخصوص صحة طفلها.
- رفع كفاءة الطاقم الطبي في المراكز الصحية عبر أداة داعمة للقرار الطبي اليومي.
- تحقيق العدالة الصحية من خلال توفير نفس مستوى التقييم للأطفال في المدن والمناطق النائية.
- توسيع قاعدة البيانات الطبية بشكل آمن لتغذية الخوارزميات وتحسين أدائها بمرور الوقت.
- تعزيز الوعي المجتمعي بأهمية الفحص المبكر والعناية بالأمراض الوراثية المزمّنة.
- دعم البحث العلمي في مجالات الصحة العامة وتحليل الوجه كوسيلة تشخيصية مستقبلية.
- تعزيز السيادة الرقمية الوطنية عبر إنشاء قاعدة بيانات صحية تُدار محليًا داخل المؤسسات الجزائرية.
- المساهمة في "حرب البيانات" من خلال تمكين الدولة من امتلاك مواردها الطبية وتوظيفها في تطوير المنظومة الصحية الوطنية.
- دعم المؤسسات الصحية في الجزائر بتقنيات ذكاء اصطناعي حديثة تُعزّز استقلاليتها الرقمية وتزيد من قوتها على المستوى المغربي والإفريقي

5.1 الجدول الزمني

1.5.1: التخطيط والتحضير

- دراسة الاحتياجات الصحية.
- تحليل السوق والمستفيدين.
- تشكيل الفريق وتحديد الشركاء الطبيين والتقنيين.

2.5.1 التصميم و التطوير

- تطوير واجهة المستخدم.
- تدريب خوارزميات الذكاء الاصطناعي.
- دمج الأنظمة واختبار أولي.

3.5.1 الإطلاق والتسويق

- تجربة ميدانية في عدد محدود من المراكز.
- حملات توعوية للأمهات والكوادر الطبية.
- شراكات مع وزارات الصحة والمؤسسات الاجتماعية.

4.5.1 الدعم والتحسين المستمر

- جمع البيانات وتحسين الأداء.
- إضافة أمراض جديدة مع مرور الوقت.
- تطوير واجهات المستخدم حسب ملاحظات الميدان

هذا الجدول الزمني

المرحلة	المدة الزمنية (الأشهر)
التخطيط و التحضير	3
التصميم والتطوير	5
الإطلاق و التسويق	3
الدعم والتحسين المستمر	مستمر
إجمالي المدة الزمنية	11 شهر

جدول 1.1: الجدول الزمني لإطلاق التطبيق

المحور 2

الجوانب الابتكارية

1.2 طبيعة الابتكارات

يُعرّف الابتكار التكنولوجي بأنه عملية تطوير وتطبيق أدوات وحلول تقنية جديدة تُحدث فرقاً ملموساً في طريقة تقديم الخدمات أو تحسين جودة الحياة. وفي هذا السياق، يقوم مشروعنا على ابتكار مزدوج يجمع بين جهاز ذكي موجه لمراكز التلقيح و تطبيق هاتفي مخصص للأولياء، وكلاهما يعتمد على تقنيات الذكاء الاصطناعي لتحليل ملامح الوجه والكشف المبكر عن بعض الاضطرابات ذات البُعد الجيني أو العصبي، مثل متلازمة داون وبعض الحالات الوراثية ذات السمات الوجهية المميزة. الجهاز صُمم ليكون أداة عملية وبمبسطة داخل المراكز الصحية، حيث يدمج كاميرا ومعالجاً صغيراً قادراً على المساعدة في التشخيص الأولي ودعم الطواقم الطبية أثناء الفحوصات الروتينية. في المقابل، يُمكن التطبيق الهاتفي الأولياء من الاستفادة من الخدمة بعد انتهاء فترة التلقيح، عبر نظام اشترك يتيح إجراء الفحص من المنزل باستخدام كاميرا الهاتف. وبذلك، يجسد المشروع ابتكاراً يربط بين البُعد الطبي والتقني من جهة، والبُعد الاجتماعي والتنظيمي من جهة أخرى، مما يعزز من قيمة الحل المقترح كأداة صحية رقمية تسهم في الكشف المبكر وتحسين التكفل بالحالات المرضية.

2.2 مجالات الابتكار

يرتكز مشروع "NeoFaceMed" على ابتكار حقيقي يجمع بين الذكاء الاصطناعي والطب من خلال جهاز ذكي مدمج مع تطبيق مرافق، يساهم في الكشف المبكر عن الأمراض الوراثية عبر تحليل ملامح الوجه. لا يقتصر الابتكار على الجانب التقني فحسب، بل يمتد أيضاً إلى نموذج أعمال مرن يتيح اعتماده على نطاق واسع داخل القطاع الصحي، مما يجعل المشروع حلاً عملياً وفعالاً لمشكلة طبية حساسة.

1.2.2 تطبيق تكنولوجيا متقدمة

يعتمد المشروع على أحدث تقنيات الذكاء الاصطناعي في تحليل الصور، ويقدم الميزات التالية:

- جهاز طبي ذكي مستقل: مزود بوحدة معالجة مدمجة، قادر على التقاط صور الوجه وتحليلها بشكل مباشر دون الحاجة إلى أجهزة إضافية.
- تطبيق هاتفي: مخصص للأولياء للاشتراك في الخدمة، يوجههم بخطوات تصوير صحيحة لضمان جودة عالية، ثم يقوم برفع الصورة للمزامنة مع الجهاز وتحليلها، مع عرض النتائج بشكل فوري وسهل الاستخدام.
- تقنيات الرؤية الحاسوبية (Computer Vision): تُستخدم لاستخلاص وتحديد السمات الوجهية التي ترتبط باحتمالية وجود أمراض معينة.
- نماذج تعلم عميق (Deep Learning Models): مدربة على مجموعات بيانات طبية متخصصة، لتعزيز دقة الكشف عن المؤشرات المرضية بناءً على الأنماط الوجهية.

2.2.2 نموذج الأعمال المبتكر

يعتمد المشروع على نموذج أعمال مرن ومتكامل يجمع بين الجهاز الذكي والتطبيق المرافق، ويتضمن:

- بيع الأجهزة للمراكز الطبية: يمكن تزويد مراكز التلقيح والعيادات بجهاز التشخيص (مزود بالكاميرا والشاشة والمعالج) مع رخصة استخدام التطبيق المرافق، مما يوفر أداة تشخيصية فورية وفعالة.
- اشتراك سنوي للجهات الصحية: يتيح للمستشفيات والمراكز الحكومية والبحثية الاستفادة من الجهاز مع التحديثات الدورية للتطبيق وخدمات الدعم الفني والتقارير السحابية.
- نسخة منزلية عبر التطبيق: موجهة للأولياء بعد سن التلقيح، عبر اشتراك شهري/سنوي في التطبيق للهاتف، لالتقاط صورة وجه الطفل وعرض التشخيص مباشرة.
- حزم شراكات مع المؤسسات: مثل الجامعات والمخابر البحثية التي يمكنها استخدام الجهاز والتطبيق لأغراض البحث والتطوير.

3.2.2 العروض والخصومات

- نسخة تجريبية مجانية: موجهة للمستخدمين الجدد بهدف تجربة النظام والتعرف على خصائصه قبل الشراء.
- رخص مؤسسية مخفضة: للمراكز الصحية الحكومية أو المؤسسات البحثية، مع إمكانية التفاوض حسب حجم الاستخدام.
- خصومات خاصة: للأطباء الشباب، الطلبة، أو المشاريع الناشئة في مجال الصحة الرقمية.
- تخصيص مرن للوظائف: يمكن تعديل بعض وظائف التطبيق حسب طبيعة كل مؤسسة (حجم البيانات، اللغة، التقارير التفصيلية...).

4.2.2 التحليل البياني

يعتمد المشروع على تحليل البيانات لدعم دقة الأداء وتحسين تجربة المستخدم، من خلال:

- تحليل بيانات الاستخدام: يتم جمع بيانات حول كيفية استعمال الجهاز والتطبيق، مثل عدد الفحوصات المنجزة، والأوقات الأكثر شيوعاً للاستعمال، لمعرفة السلوك العام وتحسين الخدمات.
- تحليل الصور: تُعالج صور وجوه الأطفال الملتقطة بالجهاز باستخدام تقنيات الذكاء الاصطناعي، بهدف كشف العلامات الصحية المرتبطة بأمراض معينة بدقة أكبر.
- تحليل التفاعلات: تتم دراسة كيفية تفاعل الأولياء أو الطاقم الطبي مع واجهة الجهاز والتطبيق، لتطويرها وجعلها أبسط وأكثر سهولة.
- تحليل الأداء العام: مثل سرعة التشخيص، ونسبة دقة النتائج مقارنة بالطرق التقليدية، لتحديد مجالات التحسين المستقبلي وضمان جودة الخدمة.

يساعد هذا التحليل البياني في تحسين دقة النماذج، وتخصيص التوصيات، وتقديم تجربة استخدام أفضل وأكثر فعالية.

5.2.2 التكامل الشامل

يهدف المشروع إلى توفير تجربة استخدام سلسة ومترابطة من خلال ربط التطبيق بأنظمة وتقنيات أخرى مستخدمة في القطاع الصحي، ويشمل ذلك:

- الربط مع أنظمة مراكز التلقيح: يمكن للجهاز تخزين البيانات بشكل محلي أو مزامنتها مع قاعدة بيانات مركزية خاصة بوزارة الصحة، لتفادي الإدخال اليدوي المتكرر وضمان متابعة دقيقة لسجلات الأطفال.
- التكامل مع التطبيق المخصص للأباء: بحيث يستطيع الأولياء الاطلاع على نتائج الفحوصات بعد كل تلقيح، والحصول على تنبيهات أو نصائح صحية مرتبطة بحالة الطفل.
- إتاحة البيانات للبحث الصحي: بشكل مجهول الهوية (دون كشف بيانات شخصية)، ما يساعد مراكز البحث والجامعات على دراسة المؤشرات الصحية للأطفال وتحسين دقة النماذج المستقبلية.

هذا التكامل يجعل الجهاز والتطبيق جزءاً فعالاً من منظومة الرعاية الصحية الرقمية، ويعزز من سهولة استخدامه وانتشاره في المراكز الصحية ومحيطها.

6.2.2 الاستدامة البيئية

يتبنى المشروع منظوراً بيئياً واعياً ينسجم مع مبادئ التنمية المستدامة في قطاع الصحة الرقمية:

- تقليل التنقلات: من خلال تمكين الكشف المبدي عبر صور الوجه، يُقلل المشروع من الحاجة إلى تنقل المرضى لمراكز التشخيص المتعددة، مما يخفف من الأثر البيئي المرتبط بوسائل النقل.
- تحول رقمي صديق للبيئة: يعتمد المشروع على العمل الرقمي الكامل، مما يُغني عن الورق والأدوات التقليدية، وبالتالي يُقلل من استهلاك الموارد ويحد من النفايات الطبية.
- بنية تقنية مستدامة: يستخدم خوادم سحابية (أو قابلة للتطبيق لاحقاً) ذات كفاءة في استهلاك الطاقة، ما يعزز من جهود التحول البيئي داخل المؤسسات الطبية.

بهذا يساهم المشروع في دعم المبادرات الصحية المستدامة والحد من الأثر البيئي للرعاية الصحية التقليدية.

المحور 3

التحليل الاستراتيجي للسوق

1.3 عرض القطاع السوقي

1.1.3 السوق المحتمل

يمثل السوق المحتمل جميع المؤسسات الصحية العامة والخاصة، خاصة تلك التي تقدم خدمات للأطفال، مثل مراكز التلقيح والمستشفيات، بالإضافة إلى العيادات المتخصصة في طب الأطفال. ومع تزايد وعي المجتمع حول أهمية التشخيص المبكر للأمراض الوراثية، والانتشار المتزايد لهذه الأمراض، فإن الحاجة إلى حل رقمي متكامل يجمع بين جهاز ذكي وتطبيق هاتفي مثل "NeoFaceMed" أصبحت ضرورية وملحة.

2.1.3 أعداد السوق المحتملين

استنادًا إلى تصريحات السيد عبد الحق سايحي، وزير الصحة الجزائري، خلال اليوم الدراسي حول ترقية صحة الأم والطفل المنعقد بالعاصمة الجزائرية سنة 2024، وبناءً على إحصائيات رسمية وتقارير وطنية، يمكن تقدير السوق المستهدف لمشروع الفحص الذكي للأطفال في الجزائر كما يلي:

- عدد الولادات السنوية في الجزائر: حوالي 895,000 ولادة سنويًا،
- عدد المراكز الصحية: 841 مركزًا صحيًا.
- عدد العيادات متعددة الخدمات الخاصة بالأمومة والطفولة: حوالي 1,841 عيادة.
- عدد المؤسسات العمومية الاستشفائية: 246 مؤسسة.
- عدد المراكز الاستشفائية الجامعية: أكثر من 16 مركزًا.
- عدد المؤسسات الاستشفائية الخاصة: أكثر من 269 مؤسسة موزعة عبر مختلف ولايات الوطن.

تُبرز هذه الإحصائيات حجم البنية التحتية الصحية الوطنية الموجهة للأمومة والطفولة، مما يعكس بيئة خصبة لتطبيق الحلول الرقمية والذكية في مجالات التشخيص المبكر والرعاية الوقائية للأطفال. كما تؤكد هذه الأرقام الإمكانيات الكبيرة التي يمكن أن يستفيد منها مشروعنا من حيث الشراكات والتكامل مع المؤسسات الصحية القائمة، وهو ما يعزز من فرص نجاح مشروع الفحص الذكي ضمن منظومة الصحة الرقمية في الجزائر.

3.1.3 السوق المستهدف

السوق المستهدف ينقسم إلى فئات محددة:

- الآباء والأمهات: الباحثين عن وسيلة ذكية وموثوقة لفحص أطفالهم في المنزل أو في المركز الصحي.
- الأطباء وأطقم التمريض: كأداة مساعدة لاتخاذ قرار طبي أولي أو دعم التشخيص السريري.
- مراكز التلقيح والرعاية الأولية: كمؤسسات تستخدم النظام كجزء من خدماتها الوقائية.
- الجهات الحكومية: خاصة وزارات الصحة والتعليم العالي، كمشترين أو داعمين لاعتماد الأداة ضمن سياسات الوقاية الوطنية

2.3 التحليل الإستراتيجي للسوق

1.2.3 من سيستخدم جهازنا؟

يستهدف الجهاز عدة فئات رئيسية من المستخدمين، تشمل:

- الفئات الأساسية:
- تُستخدم هذه التقنية داخل مراكز التلقيح أثناء الزيارات الدورية للأطفال، حيث تُمكن الطاقم الطبي من إجراء فحص ذكي وسريع للكشف المبكر عن المؤشرات الوراثية والعصبية.
- الأطباء والمرضى في مراكز التلقيح والمستشفيات: مسؤولون عن تشغيل الجهاز، التقاط الصور، وتوثيق النتائج ضمن الملفات الصحية للأطفال.
- الفئات الثانوية:
- المؤسسات الصحية والعيادات الطبية والمراكز المجاورة: يمكنها استخدام التطبيق كأداة دعم في الكشف المبكر عن بعض الحالات الوراثية والعصبية، مثل المؤشرات الوجهية، مما يساعد على تسريع التشخيص وتحسين دقته
- المستشفيات والعيادات التي تبحث عن وسائل فعالة وغير مكلفة لتعزيز أدواتها التشخيصية: خاصة في مجالات الوراثة العصبية وطب الأطفال، حيث يمكن للأطباء والباحثين والأكاديميين استخدام التطبيق في تحليل البيانات ودعم الدراسات السريرية.

2.2.3 من سيستخدم تطبيقنا؟

- الفئات الأساسية:
- الأطفال الذين تجاوزوا سن التلقيح: يمثلون المستخدمين الرئيسيين، حيث يمكن متابعة حالتهم الصحية بعد انتهاء فترة التلقيح الإجباري.
- الآباء والأمهات: يقومون بالتقاط صور لأبنائهم من المنزل وتحليلها عبر التطبيق للحصول على نتائج دورية.
- الفئات الثانوية:
- العيادات الخاصة: تعتمد التطبيق كخدمة وقائية إضافية لزيائنها.
- الأطباء وأخصائيو التشخيص عن بُعد: يراجعون نتائج التطبيق ويقدمون رأيًا طبيًا أوليًا عند الحاجة.
- الباحثون والأكاديميون: يستفيدون من البيانات بشكل مجهول في أبحاث تتعلق بالكشف المبكر عن الأمراض العصبية والوراثية.

3.2.3 دوافع الاستخدام

- الكشف المبكر: يساهم الجهاز في المراكز والتطبيق المنزلي في الكشف المبكر عن الاضطرابات العصبية والوراثية عند الأطفال، مما يقلل فترات التأخر في التشخيص.
- سهولة الوصول: التطبيق يتيح متابعة صحة الأطفال من المنزل بسهولة، والجهاز يوفر تحليلًا دقيقًا في المراكز الصحية.
- الدقة والموثوقية: يعتمد كلا النظامين على نماذج ذكاء اصطناعي متقدمة مدربة على حالات حقيقية.
- تعزيز الوقاية والخدمات الإضافية: يتيح الجهاز للعيادات تقديم خدمات وقائية، بينما يمكن للتطبيق أن يوفر متابعة صحية مرنة للأهالي حسب الحاجة.
- رفع وعي المجتمع: يعزز الوعي بأهمية التشخيص المبكر ويشجع على العناية بصحة الأطفال.

4.2.3 مبررات اختيار هذا السوق

- ارتفاع عدد المواليد: يُولد سنويًا في الجزائر حوالي 800,000 طفل، ما يشكل قاعدة كبيرة للاستفادة من أدوات التشخيص المبكر سواء عبر الجهاز أو التطبيق.
- نقص أدوات التشخيص المتقدمة: كثير من المؤسسات الصحية تفتقر إلى حلول تعتمد على الذكاء الاصطناعي، ما يفتح المجال لتبني الجهاز في المراكز والتطبيق للأهالي.

- **التحول الرقمي في القطاع الصحي:** تسعى الجزائر تدريجيًا إلى رقمنة الخدمات الصحية، ما يجعل هذا التوقيت مثاليًا لعرض حلول مبتكرة تعتمد على الجهاز والتطبيق معًا.
- **غياب المنافسة المباشرة:** لا توجد حاليًا حلول محلية مماثلة، مما يمنح المشروع فرصة زيادة السوق.
- **محدودية التجهيزات في مناطق كثيرة:** كثير من المستوصفات لا تمتلك أجهزة تشخيص دقيقة، ما يبرز دور الجهاز في المراكز والتطبيق لتوسيع الوصول للخدمات.

5.2.3 إمكانية إبرام عمليات شراء

يمثل مشروع NeoFaceMed حلاً متكاملًا يجمع بين جهاز متخصص في المراكز الصحية وتطبيق ذكي للأهالي. يتيح الجهاز للعيادات والمستشفيات تقديم خدمات وقائية وتشخيصية متقدمة، بينما يمكن التطبيق الأهالي من إجراء فحص مبكر لأبنائهم في المنزل، بغض النظر عن وجود أي أعراض معروفة، بحيث يكتشف أي مؤشرات محتملة للأمراض المستهدفة، ومن ثم التوجه للطبيب المختص عند الحاجة. تتيح آليات التسعير المرنة، سواء عبر الاشتراكات (شهري أو سنوي) أو البيع المباشر للجهاز، بالإضافة إلى الشراكات مع الجهات الحكومية والخاصة، تغطية واسعة للسوق، بما فيها المناطق ذات الموارد المحدودة. كما يضمن التصنيع المحلي أو شبه المحلي للجهاز تقديم أسعار تنافسية وزيادة اعتماد المشروع كأداة قياسية في الرعاية الصحية الأولية، مما يجعل NeoFaceMed خيارًا جذابًا للمؤسسات الصحية والأهالي على حد سواء.

3.3 قياس شدة المنافسة في السوق الجزائرية

1.3.3 المنافسون المباشرون

يُعد مشروع NeoFaceMed الأول من نوعه في الجزائر من حيث توظيف الذكاء الاصطناعي لتحليل ملامح الوجه للكشف المبكر عن الأمراض الوراثية والعصبية. لا توجد حلول محلية مشابهة في السوق تعتمد هذا النوع من التشخيص غير المباشر، سواء من قبل مؤسسات طبية أو شركات تكنولوجيا صحية. هذا الغياب التام للمنافسين المباشرين يمنح المشروع ميزة تنافسية واضحة وفرصة لقيادة السوق في هذا المجال المتخصص.

2.3.3 المنافسون غير المباشرين

رغم عدم وجود حلول محلية تعتمد على تحليل ملامح الوجه بالذكاء الاصطناعي، إلا أن هناك بعض الأدوات غير المباشرة التي تقدم خدمات مشابهة في الأهداف، لكنها تختلف تمامًا في آلية العمل. من بين هذه الأدوات:

- **Face2Gene:** منصة عالمية تستخدم الذكاء الاصطناعي لتحليل صور الوجه بهدف دعم تشخيص المتلازمات الوراثية، لكنها تستهدف بشكل رئيسي الأسواق الأجنبية، ولا توفر دعمًا مخصصًا للسياق الطبي المحلي الجزائري.
- **التحاليل الجينية التقليدية:** مثل فحوصات DNA والفحوصات المخبرية للكشف عن الأمراض الوراثية، وهي أدوات دقيقة لكنها تتطلب وقتًا أطول وتكلفة أعلى، ولا تتيح نتائج فورية كما هو الحال مع NeoFaceMed.

• الخبرة الطبية التقليدية: تعتمد على الفحص السريري والتقدير البشري للملامح الوجهية، لكنها تفتقر إلى القدرة على التحليل الدقيق والآلي الذي يقدمه الذكاء الاصطناعي.

يمثل مشروع "NeoFaceMed" نظامًا ذكيًا متكاملًا يجمع بين تطبيق للهاتف الذكي وجهاز مخصص للمراكز الصحية، بهدف تسهيل التشخيص المبكر باستخدام الصور فقط. يمنح هذا الجمع بين التطبيق والجهاز ميزة فريدة في السوق مقارنة بالبدائل المتاحة، حيث يتيح الجهاز الصغير الحجم المزود بكاميرا ومعالج وشاشة استخدامه داخل المراكز الصحية أو نقله إلى المناطق الريفية، ما يعزز قدرة المؤسسات الصحية على تقديم خدمات وقائية دقيقة وفعالة حتى في البيئات محدودة الإمكانيات.

3.3.3 التحليل العام للمنافسة

يُعد مشروع NeoFaceMed مبادرة مبتكرة في سوق ناشئ، حيث يجمع بين جهاز متكامل للمراكز الصحية وتطبيق للهاتف الذكي يعتمد على تحليل ملامح الوجه بالذكاء الاصطناعي لأغراض طبية، خاصة في الجزائر. ونظرًا لحدائثة الفكرة وغياب المنافسين المحليين المباشرين، يحتل المشروع موقعًا رياديًا واضحًا في هذا المجال، مع تنامي الاهتمام بالتحول الرقمي في القطاع الصحي.

هذا التميز يمنح المشروع فرصة قوية للانتشار السريع، وتأسيس ثقة مبكرة لدى المؤسسات الطبية، من خلال تقديم قيمة مضافة عملية تتمثل في السرعة والدقة وسهولة الاستخدام مقارنة بالطرق التقليدية أو البدائل المكلفة. كما أن طبيعة التطبيق القابلة للاستخدام على الهاتف الذكي، موازاةً مع قدرة الجهاز على العمل داخل المراكز الصحية، تعزز من قدرة المشروع على التكيف مع احتياجات السوق المحلي بمختلف مناطقه، بما فيها المناطق الريفية أو ضعيفة البنية التحتية. بفضل هذا السبق، يمكن للمشروع أن يرسخ اسمه كمرجع أول في هذا المجال، مما يصعب على المنافسين لاحقًا الدخول بنفس القوة أو التأثير، ويؤكد ريادة المشروع في دمج التكنولوجيا الذكية مع الرعاية الصحية الأولية.

4.3 الإستراتيجية التسويقية

1.4.3 تحليل السوق والفهم العميق للزبائن

يعتمد تحليل السوق على فهم الفئة المستهدفة بدقة، وهم الأطباء، مدراء المؤسسات الصحية، وأولياء الأطفال الذين يعانون من أعراض يصعب تشخيصها مبكرًا. هؤلاء الزبائن يبحثون عن أدوات فعالة وسريعة تساعدهم في اتخاذ قرارات دقيقة دون الحاجة إلى فحوصات مكلفة أو غير متوفرة محليًا. كما تبرز أهمية فهم التباين بين المناطق الحضرية والريفية في الوصول إلى الخدمات الصحية.

2.4.3 تحديد الميزة التنافسية

الميزة الأساسية لـ NeoFaceMed تكمن في كونه أول نظام محلي يعتمد على الذكاء الاصطناعي لتحليل ملامح الوجه للكشف عن الأمراض، إلى جانب ذلك:

- سهولة الاستخدام من طرف غير المتخصصين.
- القدرة على التشخيص السريع.
- جهاز منخفض التكلفة في المستقبل مقارنة بالحلول التقليدية أو الأجنبية.
- احترام خصوصية المستخدمين وتخزين البيانات محليًا.
- الإسهام في تعزيز الأمن القومي من خلال حماية أمن المعلومات الصحية وتطوير قاعدة بيانات وطنية آمنة ومستقلة.

3.4.3 استخدام قنوات التسويق الفعالة

سيتم اعتماد مزيج من القنوات لتوسيع الوصول إلى السوق:

- حملات توعوية رقمية عبر مواقع التواصل.
- ندوات طبية وورش عمل ميدانية.
- التواصل المباشر مع مدراء المستشفيات والمراكز الصحية.
- مقالات علمية في مجلات طبية.
- دعم من وزارة الصحة لتسهيل التبني المؤسسي.

4.4.3 تحسين تجربة المستخدم

لضمان استخدام سلس وفعال:

- تصميم واجهة بسيطة وواضحة.
- تضمين شروحات مرئية ونصية داخل النظام.
- تطوير خاصية التفاعل الذكي مع الطبيب (اقتراحات وتوصيات).

5.4.3 استراتيجيات الاحتفاظ بالعملاء

لضمان استمرارية استخدام التطبيق والجهاز وتعزيز ولاء المستخدمين:

- تقديم خطط اشتراك مرنة تلبي احتياجات الأسر والعيادات على حد سواء.
- تحديثات دورية للتطبيق والجهاز تشمل تحسينات الأداء ووظائف جديدة مستندة إلى التغذية الراجعة للمستخدمين.
- توفير محتوى علمي وتوعوي دوري لتعزيز معرفة المستخدمين وفائدتهم العملية.

6.4.3 الشراكات الاستراتيجية

سيتم بناء شبكة شراكات مع:

- مستشفيات جامعية ومراكز طب الأطفال.
- معاهد تكوين شبه طبي.
- شركات توزيع الأجهزة الطبية لتجهيز العيادات.
- جامعات ومخابر بحث لدعم تطوير الخوارزميات.

7.4.3 تقييم الأداء والتحسين المستمر

تعتمد الخطة على مراقبة مؤشرات أداء رئيسية:

- عدد المستخدمين الفعليين.
- نسبة الدقة في التشخيص مقارنة بالتشخيص الطبي التقليدي.
- معدل الاستخدام الشهري.
- نسبة الرضا من خلال استبيانات دورية.

كما سيتم تحليل كل هذه البيانات دورياً وتحديث الاستراتيجية بناءً عليها لتحقيق أفضل نتائج ممكنة.

المحور 4

خطة الإنتاج والتنظيم

1.4 خطة الإنتاج

1.1.4 التخطيط والتحضير

- تحديد أهداف المشروع التقنية والطبية لكل من التطبيق والجهاز الطبي، مع التركيز على قدرة الجهاز على إجراء التحاليل بسرعة ودقة داخل المراكز الصحية.
- جمع بيانات موثوقة حول ملامح الوجه والأمراض الوراثية من مراكز طبية ومستشفيات لتدريب النماذج الذكية للجهاز والتطبيق.
- اختيار الأدوات والخوارزميات الملائمة لتطبيق التعلم الآلي في كل من التطبيق والهاردوير لضمان أعلى دقة تشخيصية.
- تشكيل فريق متعدد التخصصات يشمل مطورين، خبراء ذكاء اصطناعي، وأطباء، لضمان تكامل النظامين.
- إعداد ميزانية لتغطية تكاليف تطوير الجهاز والتطبيق، بما في ذلك المواد والمكونات المادية للجهاز والبرمجيات.

2.1.4 التصميم والتطوير

- تصميم واجهة المستخدم للتطبيق بطريقة بسيطة وواضحة، مع مراعاة سهولة التعامل مع الجهاز داخل المراكز الصحية.
- تطوير الجهاز ليكون مدمجًا يحتوي على كاميرا، معالج، وشاشة، قادرًا على إجراء تحاليل مباشرة ودقيقة للوجه.
- تدريب النماذج الذكية على قواعد بيانات تحتوي على صور لحالات مؤكدة، وضمان دقتها في التطبيق والجهاز على حد سواء.

- اختبار تكامل الجهاز مع التطبيق بحيث يمكن للأهالي استخدام التطبيق في المنزل، بينما يُستخدم الجهاز داخل المراكز الصحية لمراقبة الحالات الأكثر تفصيلاً.

3.1.4 الإطلاق والتسويق

- اختيار مناطق تجريبية لإطلاق النظام، مع التركيز على مراكز التلقيح والعيادات التي ستستخدم الجهاز بشكل أساسي.
- تنظيم ورش عمل لتدريب الأطباء والمستخدمين على استخدام الجهاز والتطبيق معاً لضمان الاستفادة القصوى.
- تسليط الضوء على تميز الجهاز بقدرته على تشخيص سريع ودقيق، مما يعزز ثقة المستخدمين وسمعة المشروع في السوق.

4.1.4 الدعم والتحسين المستمر

- توفير قنوات دعم فني فوري (هاتف، واتساب، شات داخل التطبيق).
- اعتماد سياسة "التحسين المستمر" بتحديث الخوارزميات عند توفر بيانات جديدة.
- إنشاء قاعدة بيانات مركزية تخزن نتائج الاستخدام لتحسين النظام لاحقاً.

5.1.4 التقييم والاستدامة

- تحليل أداء النظام باستخدام مؤشرات دقيقة مثل معدل الاكتشاف المبكر وعدد المستخدمين الفعليين للتطبيق والجهاز.
- إعداد تقارير تقييم شهرية لمتابعة فعالية النظام وتحسين العمليات.
- وضع خطة لضمان الاستدامة المالية، عبر نماذج اشتراك مرنة ودعم حكومي أو شراكات مع مؤسسات صحية.
- دراسة فرص التوسع مستقبلاً إلى أسواق محلية أخرى أو دولية، مع التركيز على المراكز الصحية والتطبيقات المنزلية.
- تقييم رضا المستخدمين لضمان التحسين المستمر وسهولة التكيف مع احتياجات الأهالي والأطباء.

2.4 الجدول الزمني

المرحلة	النشاط	المدة الزمنية (الأسابيع)
التخطيط و التحضير	تحديد المتطلبات و جمع البيانات	10
التصميم والتطوير	تصميم قاعدة البيانات، تطوير النموذج الأولي، برمجة التطبيق، تصميم واجهة المستخدم، تطوير خوارزميات الذكاء الاصطناعي، اختبار التطبيق	20
الإطلاق و التسويق	إعداد استراتيجية التسويق، إطلاق النسخة التجريبية، إطلاق النسخة النهائية، تنفيذ حملة تسويقية	8
الدعم والتحسين المستمر	دعم العملاء، جمع الملاحظات، التحديثات الدورية	مستمر
التقييم و الإستدامة	تقييم الأداء، استراتيجيات الإستدامة	مستمر
إجمالي المدة الزمنية	//	38 أسبوع

جدول 1.4: الجدول الزمني التفصيلي لإطلاق التطبيق

المحور 5

الخطة التسويقية

1.5 الخطة المالية للمشروع ونموذج العمل التجاري

”NeoFaceMed” هو برنامج ذكي يندرج ضمن الابتكار الاجتماعي الرقمي، يستخدم الذكاء الاصطناعي لتحليل ملامح وجه الأطفال للكشف المبكر عن الأمراض الوراثية. المشروع يتماشى مع أهداف الدولة الجزائرية في رقمنة قطاع الصحة وتقليص الفجوة الصحية بين المناطق، ويستهدف مراكز التلقيح، المستوصفات، وعيادات الأطفال. نموذج العمل قائم على ترخيص سنوي منخفض التكلفة، يضمن سهولة التبني. من أبرز مميزات المشروع:

- لا حاجة إلى مقر إداري كبير، فقط فضاء بسيط لتخزين الأجهزة، بينما يعمل الفريق عن بُعد.
- لا يتطلب توظيف دائم، إذ يعتمد على الكفاءة الذاتية لمطوريه.
- يعتمد على خوادم (Servers) وقاعدة بيانات لتخزين الصور والنتائج بشكل آمن، دون الحاجة إلى تجهيزات باهظة أو خوادم مركزية خاصة، مما يخفض التكلفة الإجمالية ويسهل توسيع الحل عند نمو عدد المستخدمين.

1.1.5 التكاليف والرسوم

1. تكاليف لوجيستية:

تمثل النفقات الأساسية المرتبطة بتجهيز الفضاء المادي والوسائل الضرورية لضمان انطلاق المشروع:

- إيجار فضاء للتخزين: المشروع لا يحتاج إلى مقر إداري كبير، وإنما فقط فضاء بسيط لتخزين الأجهزة والمكونات، بتكلفة سنوية تُقدر بـ 120,000 دج.
- تجهيزات أساسية: مكتب بسيط، رفوف لحفظ الأجهزة، طابعة، وبعض اللوازم المكتبية، بتكلفة إجمالية تُقدر بـ 80,000 دج.
- اشتراك إنترنت عالي السرعة: ضروري لضمان عمل الفريق عن بُعد ومزامنة البيانات مع الخوادم، بتكلفة سنوية تُقدر بـ 36,000 دج.

إجمالي التكاليف اللوجستية: 236,000 دج/سنة

2. تكاليف التسويق والدعاية:

تُعتبر حاسمة لتوسيع نطاق الحل والوصول إلى الفئات المستهدفة من مراكز التلقيح والأولياء:

- حملات تسويق رقمي: عبر وسائل التواصل الاجتماعي (فيسبوك، إنستغرام، تيك توك) للوصول إلى أكبر عدد ممكن من الأطباء والأولياء، بتكلفة سنوية تُقدَّر بـ 300,000 دج.
- إعداد مواد ترويجية: تشمل فيديوهات تعريفية بالجهاز والتطبيق، صور توضيحية، وإنشاء محتوى رقمي توعوي، بتكلفة تُقدَّر بـ 80,000 دج.
- تصميم الهوية البصرية: تشمل الشعار، الخطوط، الألوان الرسمية، والقوالب المستخدمة في الحملات، بتكلفة إجمالية تُقدَّر بـ 100,000 دج.

إجمالي تكاليف التسويق والدعاية: 480,000 دج/سنة

3. تكاليف التشغيل والصيانة:

تضمن استمرارية الخدمة وجودتها على المدى الطويل:

- خوادم وقاعدة بيانات: لتخزين صور ونتائج التشخيص بشكل آمن، مع إمكانية التوسع السحابي، بتكلفة سنوية تُقدَّر بـ 120,000 دج.
- الصيانة والتحديثات البرمجية: تشمل تحديث التطبيق الهاتفي والبرمجيات المدمجة في الجهاز، لضمان الجودة والامتثال للمعايير الصحية، بتكلفة تُقدَّر بـ 150,000 دج/سنة.
- الدعم التقني وخدمة الزبائن: توفير قنوات دعم عن بُعد للمراكز الصحية والمستعملين، بتكلفة سنوية تُقدَّر بـ 60,000 دج.

إجمالي تكاليف التشغيل والصيانة: 330,000 دج/سنة

4. تكاليف الإدارة:

مرتبطة بتسيير المشروع والفريق المؤسس:

- مكافآت رمزية للمؤسسين: (2 أشخاص × 25,000 دج/شهر) أي ما يعادل 600,000 دج/سنة.
- تنقلات واجتماعات ميدانية: تخص لقاءات مع مؤسسات صحية أو شركات في معارض تقنية، بتكلفة سنوية تُقدَّر بـ 60,000 دج.
- تجهيزات مكتبية إضافية: مثل الحواسيب الشخصية، الأدوات التقنية الخفيفة، والبرامج، بتكلفة تُقدَّر بـ 40,000 دج.

إجمالي تكاليف الإدارة: 700,000 دج/سنة

5. التكاليف القانونية:

ضرورة لضمان شرعية المشروع وحماية حقوق الملكية الفكرية:

- رسوم تسجيل الشركة والسجل التجاري: تُقدَّر بـ 50,000 دج (مرة واحدة).
- اشتراكات الضمان الاجتماعي (CNAS) أو (CASNOS) للمؤسسين (2 أفراد) تُقدَّر بـ 80,000 دج/سنة.
- استشارات قانونية ومحاسبية: لمتابعة العقود، الاتفاقيات، والجانب الجبائي، بتكلفة سنوية تُقدَّر بـ 60,000 دج

إجمالي التكاليف القانونية: 190,000 دج/سنة

2.5 تمويل المشروع

سيتم تمويل المشروع من قبل صندوق الشركات الناشئة الجزائري

سداد الاعتمادات والقروض سيتم تنفيذها على المدى الطويل والذي سيتم التفاوض عليه مع المنظمة المعنية.

الشرط	التكلفة
تطوير البرمجيات والتصميم	600,000
تصميم وتصنيع الجهاز الذكي (النموذج الأولي)	120,000
إنتاج أولي (10 أجهزة للتجارب الميدانية)	800,000
حملات التسويق الرقمي	200,000
إعداد المواد الترويجية	30,000
العلاقات العامة والترويج	50,000
أجهزة التشغيل والصيانة	200,000
تجهيزات مكتبية إضافية	80,000
التشغيل اليومي	60,000
مكافآت رمزية للمؤسسين	600,000
تنقلات ومهام ميدانية	60,000
تكاليف قانونية	50,000
أعباء صندوق التأمين CNAS	80,000
إيجار فضاء صغير للتخزين	120,000
الإجمالي	2,960,000

جدول 1.5: جدول التكاليف والإستثمار (التوقعات)

3.5 رقم الأعمال

1.3.5 النظرة التفاولية

ترتكز النظرة التفاولية على توقع تحقيق المشروع لنمو متسارع بفضل دمج الجهاز الذكي والتطبيق الهاتفي ضمن منظومة رقمية مبتكرة. هذا السيناريو يأخذ بعين الاعتبار قابلية الحل للاعتماد السريع في الميدان، خاصة داخل مراكز التلقيح والمستوصفات، مع زيادة تدريجية في عدد الاشتراكات من طرف الأولياء والمؤسسات الصحية.

يُفترض في هذا الإطار أن يبقى سعر الاشتراك في متناول المستخدمين دون تغيير كبير، مما يسهل عملية التنبؤ الواسع، بينما يرتفع حجم المبيعات والإيرادات بشكل تدريجي سنة بعد أخرى. ويعكس هذا التصور قدرة المشروع على تحقيق توازن مالي مستدام، من خلال الاعتماد على مصدرين رئيسيين للإيرادات: بيع الأجهزة الذكية و الاشتراكات الرقمية عبر التطبيق.

الجدولين الموالين يوضحان التطور المتوقع للإيرادات على مدى خمس سنوات، مع التمييز بين مداخيل الأجهزة ومداخيل التطبيق.

جدول 2.5: تطور الإيرادات المتوقعة من بيع الجهاز الذكي (النظرة التفاولية)

السنة	عدد الأجهزة المباعة	سعر الجهاز (دج)	الإيرادات الإجمالية (دج)
N	200	50,000	10,000,000
N+1	400	50,000	20,000,000
N+2	600	50,000	30,000,000
N+3	800	50,000	40,000,000
N+4	1000	50,000	50,000,000

جدول 3.5: تطور الإيرادات المتوقعة من التطبيق الهاتفي (النظرة التفاولية)

السنة	عدد الاشتراكات	سعر الاشتراك السنوي (دج)	الإيرادات الإجمالية (دج)
N	12,250	500	6,125,000
N+1	13,200	500	6,600,000
N+2	14,000	500	7,000,000
N+3	16,000	500	8,000,000
N+4	18,000	500	9,000,000

2.3.5 النظرة التشارؤية

يبين الجدول التالي التوقعات المالية وفق النظرة التشارؤية، حيث جرى افتراض بطء في توسع قاعدة المستخدمين وضعف وتيرة المبيعات. الهدف من هذا العرض هو إبراز الحد الأدنى الممكن من الأداء المالي للمشروع في حال واجه تحديات تسويقية أو تشغيلية، مع الحفاظ على استمرارية النشاط وقدرته على تغطية التكاليف الأساسية، سواء بالنسبة للجهاز أو للتطبيق.

السنة	عدد الأجهزة المباعة	سعر الجهاز (دج)	الإيرادات الإجمالية (دج)
N	40	50,000	2,000,000
N+1	80	50,000	4,000,000
N+2	120	50,000	6,000,000
N+3	160	50,000	8,000,000
N+4	200	50,000	10,000,000

جدول 4.5: تطور المبيعات والإيرادات للجهاز – النظرة التشارؤية

السنة	عدد الاشتراكات	سعر الاشتراك (دج)	الإيرادات الإجمالية (دج)
N	2,450	500	1,225,000
N+1	2,640	500	1,320,000
N+2	2,800	500	1,400,000
N+3	3,200	500	1,800,000
N+4	4,000	500	2,000,000

جدول 5.5: تطور الاشتراكات والإيرادات للتطبيق – النظرة التشارؤية

4.5 جداول حسابات النتائج

المصاريف / الأشهر	1	2	3	4	5	6
تطوير التطبيق	30,000	30,000	30,000	30,000	30,000	30,000
تصنيع الجهاز	50,000	50,000	50,000	50,000	50,000	50,000
تكاليف التسويق والدعاية	40,000	40,000	40,000	40,000	40,000	40,000
معدات المكتبة	/	/	/	/	/	/
صندوق أعباء التأمين CNAS	6,300	6,300	6,300	6,300	6,300	6,300
تأمين المعدات	8,000	8,000	8,000	8,000	8,000	8,000
التكاليف الإدارية	4,166	4,166	4,166	4,166	4,166	4,166
رواتب العمال	70,000	70,000	70,000	70,000	70,000	70,000
تكلفة الإيجار	40,000	40,000	40,000	40,000	40,000	40,000
الرسوم الخارجية	37,500	37,500	37,500	37,500	37,500	37,500
المحاسبون القانونيون	6,830	6,830	6,830	6,830	6,830	6,830
مجموع النفقات	292,796	292,796	292,796	292,796	292,796	292,796

جدول 6.5: جدول المصاريف (الأشهر من 1 إلى 6)

المصاريف / الأشهر	7	8	9	10	11	12	المجموع
تطوير التطبيق	30,000	30,000	30,000	30,000	30,000	30,000	360,000
تصنيع الجهاز	50,000	50,000	50,000	50,000	50,000	50,000	600,000
تكاليف التسويق والدعاية	40,000	40,000	40,000	40,000	40,000	40,000	480,000
معدات المكتبة	/	/	/	/	/	/	400,000
صندوق أعباء التأمين CNAS	6,300	6,300	6,300	6,300	6,300	6,300	75,600
تأمين المعدات	8,000	8,000	8,000	8,000	8,000	8,000	96,000
التكاليف الإدارية	4,166	4,166	4,166	4,166	4,166	4,166	49,992
رواتب العمال	70,000	70,000	70,000	70,000	70,000	70,000	840,000
تكلفة الإيجار	40,000	40,000	40,000	40,000	40,000	40,000	480,000
الرسوم الخارجية	37,500	37,500	37,500	37,500	37,500	37,500	450,000
المحاسبون القانونيون	6,830	6,830	6,830	6,830	6,830	6,830	81,960
مجموع النفقات	292,796	292,796	292,796	292,796	292,796	292,796	3,513,552

جدول 7.5: جدول المصاريف (الأشهر من 7 إلى 12)

5.5 خطة الخزينة

خطة الخزينة التي سينتهجها مشروعنا هي كالتالي في السنوات المقبلة موضحة في الجداول الموالية:

البند	01	02	03	04	05	06
المبيعات الإجمالية	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000
صافي المبيعات	1,900,000	1,900,000	1,900,000	1,900,000	1,900,000	1,900,000
المصاريف	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
هامش الربح	900,000	900,000	900,000	900,000	900,000	900,000
الربح التشغيلي	800,000	800,000	800,000	800,000	800,000	800,000
إيرادات فوائد	0	0	0	0	0	0
الربح قبل الضريبة	800,000	800,000	800,000	800,000	800,000	800,000
رسوم ضريبية	0	0	0	0	0	0
صافي الإيرادات	800,000	800,000	800,000	800,000	800,000	800,000

جدول 8.5: الإيرادات والنفقات المتوقعة خلال السنة الأولى (الأشهر 1 إلى 6)

البند	07	08	09	10	11	12
المبيعات الإجمالية	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000
صافي المبيعات	1,900,000	1,900,000	1,900,000	1,900,000	1,900,000	1,900,000
المصاريف	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
هامش الربح	900,000	900,000	900,000	900,000	900,000	900,000
الربح التشغيلي	800,000	800,000	800,000	800,000	800,000	800,000
إيرادات فوائد	0	0	0	0	0	0
الربح قبل الضريبة	800,000	800,000	800,000	800,000	800,000	800,000
رسوم ضريبية	0	0	0	0	0	0
صافي الإيرادات	800,000	800,000	800,000	800,000	800,000	800,000

جدول 9.5: الإيرادات والنفقات المتوقعة خلال السنة الأولى (الأشهر 7 إلى 12)

06	05	04	03	02	01	البند
2,400,000	2,400,000	2,400,000	2,400,000	2,400,000	2,400,000	المبيعات الإجمالية
2,280,000	2,280,000	2,280,000	2,280,000	2,280,000	2,280,000	صافي المبيعات
1,200,000	1,200,000	1,200,000	1,200,000	1,200,000	1,200,000	المصاريف
1,080,000	1,080,000	1,080,000	1,080,000	1,080,000	1,080,000	هامش الربح
960,000	960,000	960,000	960,000	960,000	960,000	الربح التشغيلي
0	0	0	0	0	0	إيرادات فوائد
960,000	960,000	960,000	960,000	960,000	960,000	الربح قبل الضريبة
0	0	0	0	0	0	رسوم ضريبية
960,000	960,000	960,000	960,000	960,000	960,000	صافي الإيرادات

جدول 10.5: الإيرادات والنفقات المتوقعة خلال السنة الخامسة (الأشهر 1 إلى 6)

12	11	10	09	08	07	البند
2,400,000	2,400,000	2,400,000	2,400,000	2,400,000	2,400,000	المبيعات الإجمالية
2,280,000	2,280,000	2,280,000	2,280,000	2,280,000	2,280,000	صافي المبيعات
1,200,000	1,200,000	1,200,000	1,200,000	1,200,000	1,200,000	المصاريف
1,080,000	1,080,000	1,080,000	1,080,000	1,080,000	1,080,000	هامش الربح
960,000	960,000	960,000	960,000	960,000	960,000	الربح التشغيلي
0	0	0	0	0	0	إيرادات فوائد
960,000	960,000	960,000	960,000	960,000	960,000	الربح قبل الضريبة
0	0	0	0	0	0	رسوم ضريبية
960,000	960,000	960,000	960,000	960,000	960,000	صافي الإيرادات

جدول 11.5: الإيرادات والنفقات المتوقعة خلال السنة الخامسة (الأشهر 7 إلى 12)

المحور 6

النموذج التجريبي الأولي

1.6 التطبيق

1.1.6 صفحة البدء

14:57

21%   



Neo Face Med

Smart Care for Your Child's Health

تُعرض هذه الصفحة عند فتح التطبيق لأول مرة، وتحتوي على شعار NeoFaceMed، مما يمنح انطباعًا أوليًا احترافيًا ويهيئ المستخدم للدخول إلى واجهة التطبيق.

2.1.6 صفحة اختيار اللغة



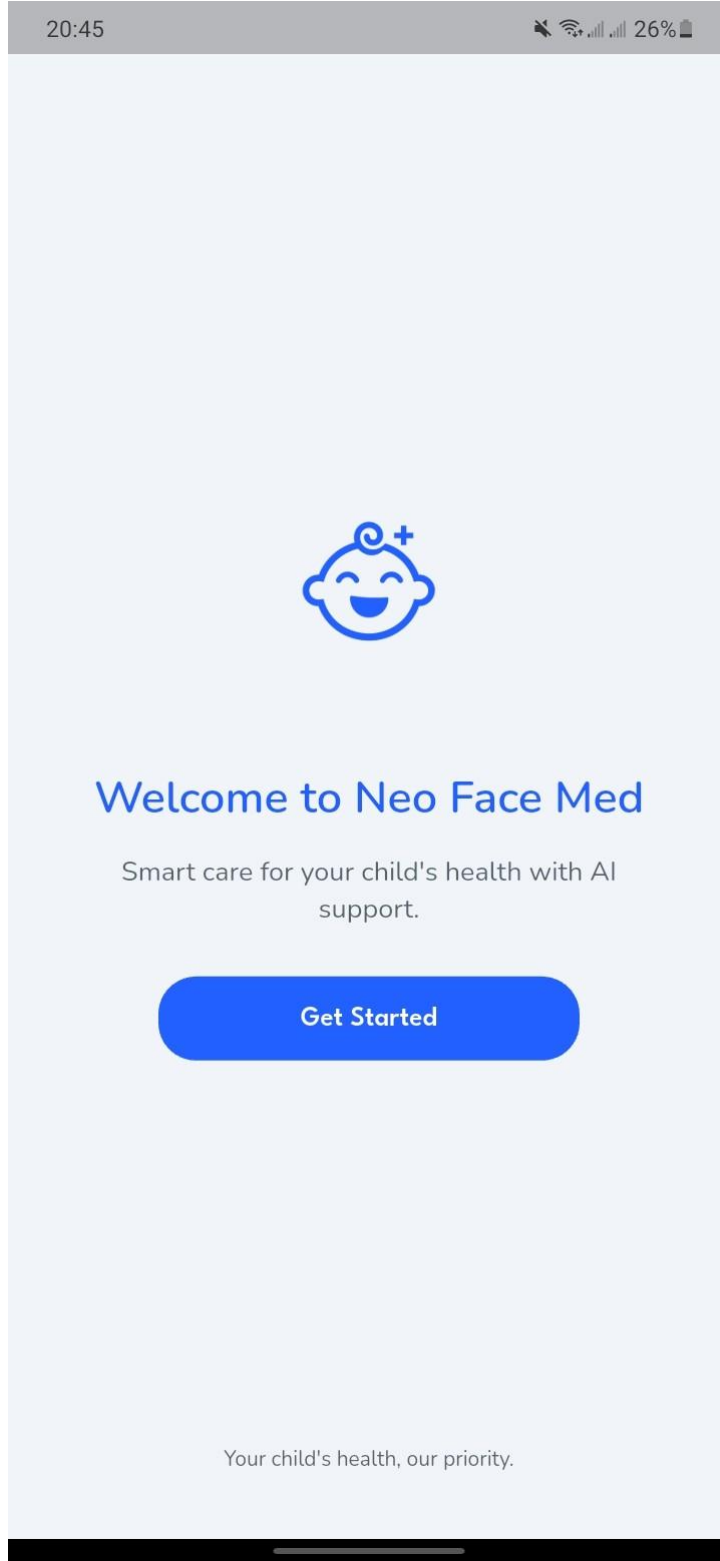
Choose Your Language



شكل 2.6: صفحة اختيار اللغة

تتيح للمستخدم اختيار لغة الاستخدام بين العربية والإنجليزية، بما يضمن تجربة سلسة ومناسبة لجميع الفئات.

3.1.6 صفحة الترحيب



شكل 3.6: صفحة الترحيب

تُقدّم هذه الصفحة ترحيبًا بالمستخدم مع نبذة مختصرة عن فكرة التطبيق وأهدافه، إضافة إلى زر ابد للانتقال إلى الخطوة التالية.

4.1.6 صفحة التسجيل

20:45 26%



Email

Password

Login

Don't have an account? [Sign Up](#)

شكل 4.6: واجهة إنشاء الحساب

5.1.6 إنشاء الحساب



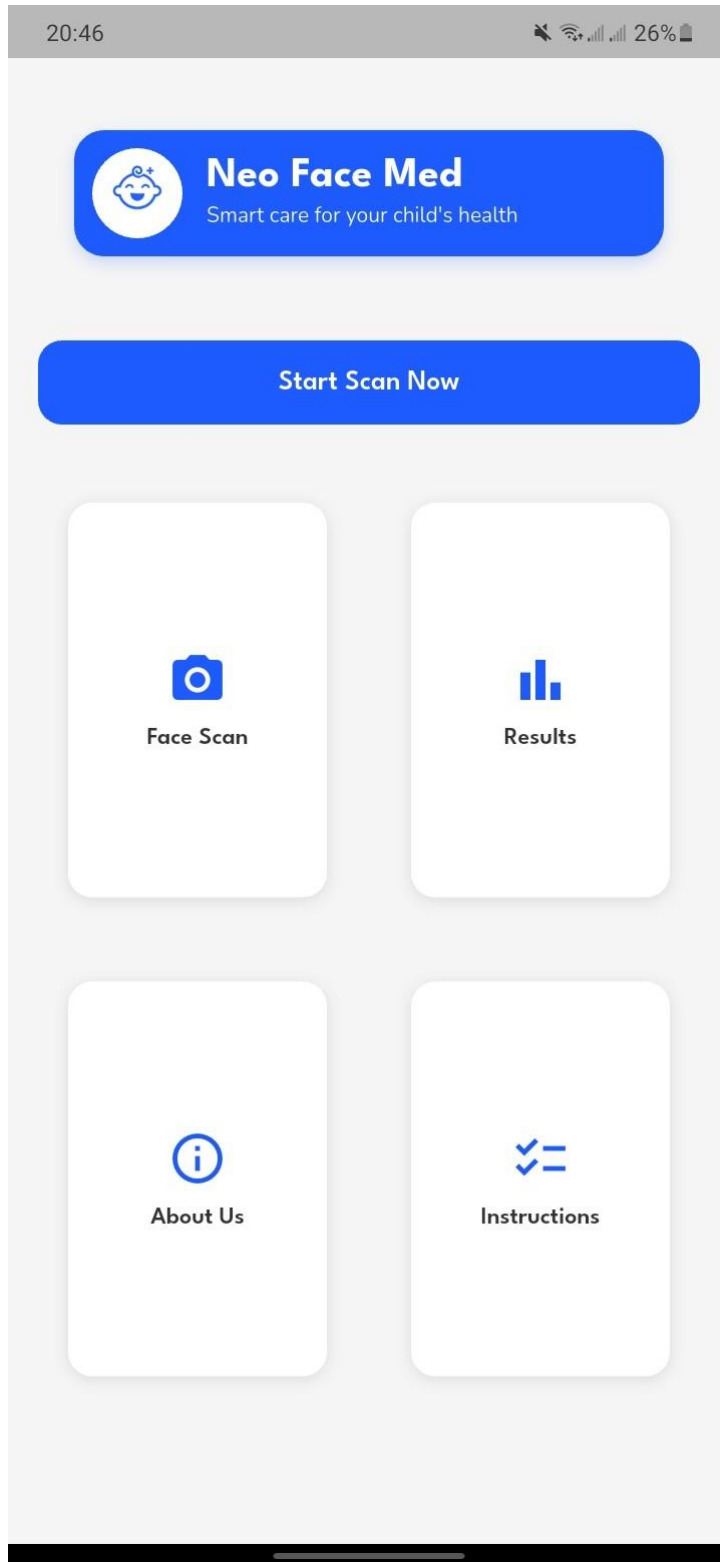
Sign Up



Already have an account? [Log In](#)

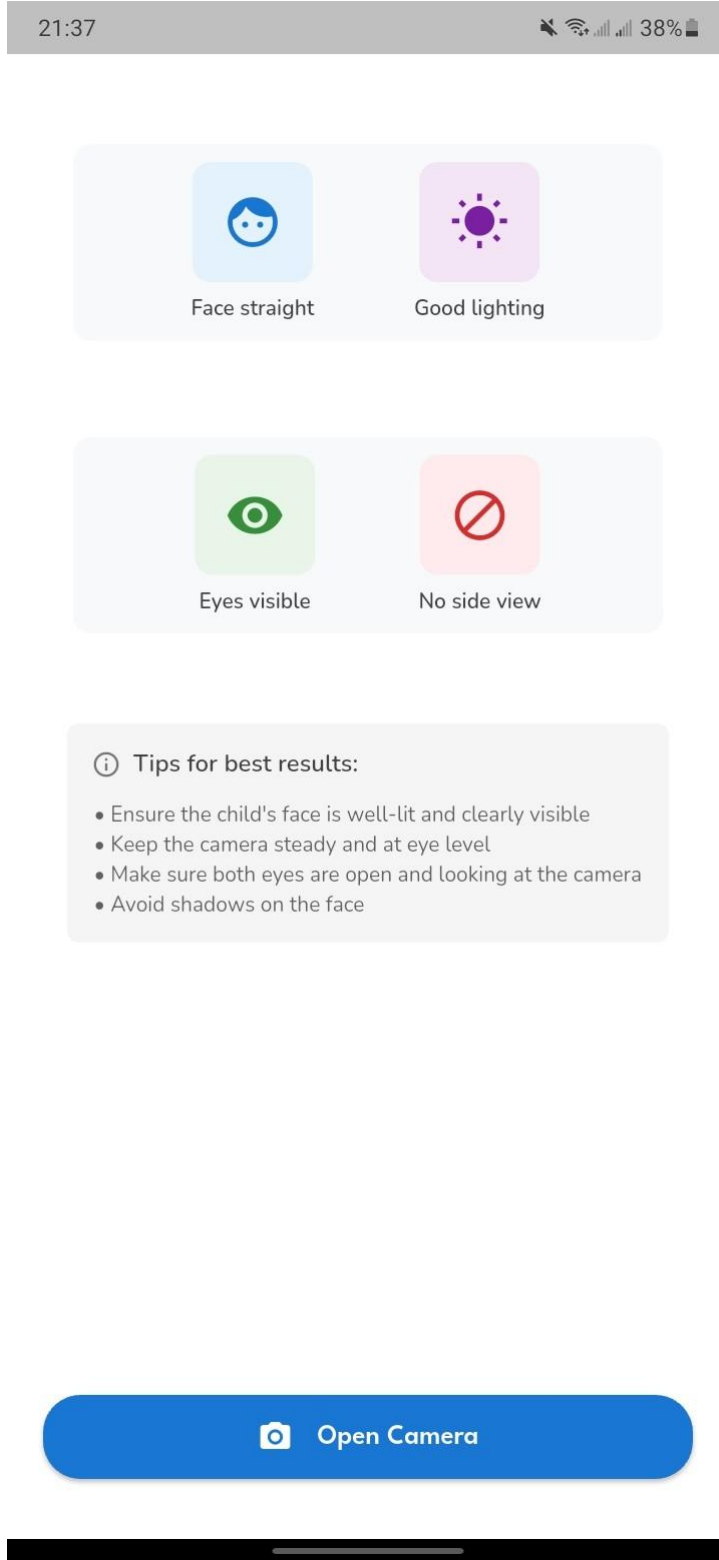
شكل 5.6: واجهة إنشاء الحساب

6.1.6 الصفحة الرئيسية



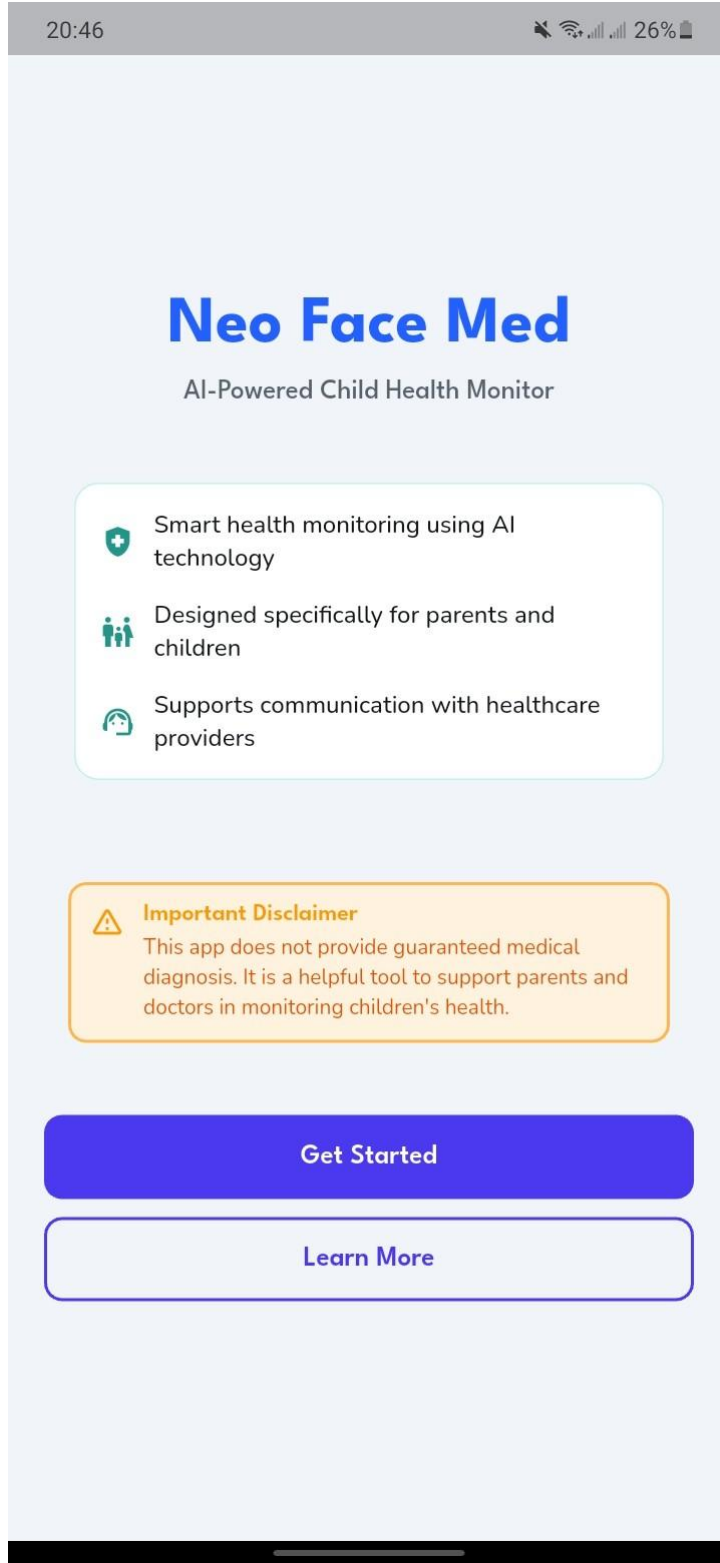
شكل 6.6: الصفحة الرئيسية

7.1.6 صفحة التعليمات



شكل 7.6: صفحة التعليمات
تحتوي على مجموعة من التعليمات المفصلة حول كيفية التقاط الصورة بالشكل الصحيح، لضمان جودة البيانات ودقة النتائج الصادرة عن النظام.

8.1.6 صفحة من نحن



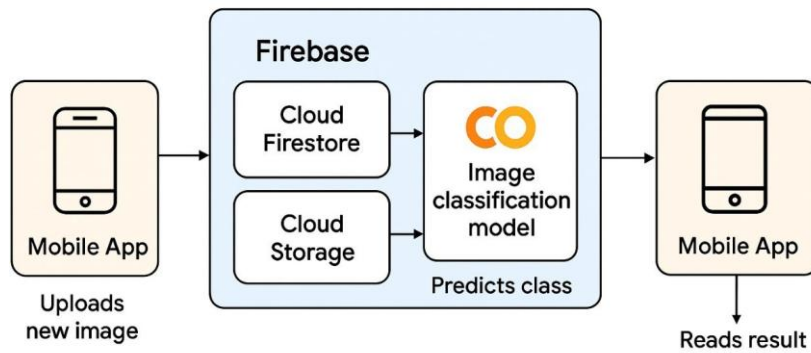
شكل 8.6: صفحة من نحن

تُعرّف بالمشروع وفريق العمل، وتعرض رؤية NeoFaceMed ورسالتها في دعم المجال الصحي من خلال توظيف الذكاء الاصطناعي لخدمة الطفل.



شكل 9.6: اكتشاف المزيد

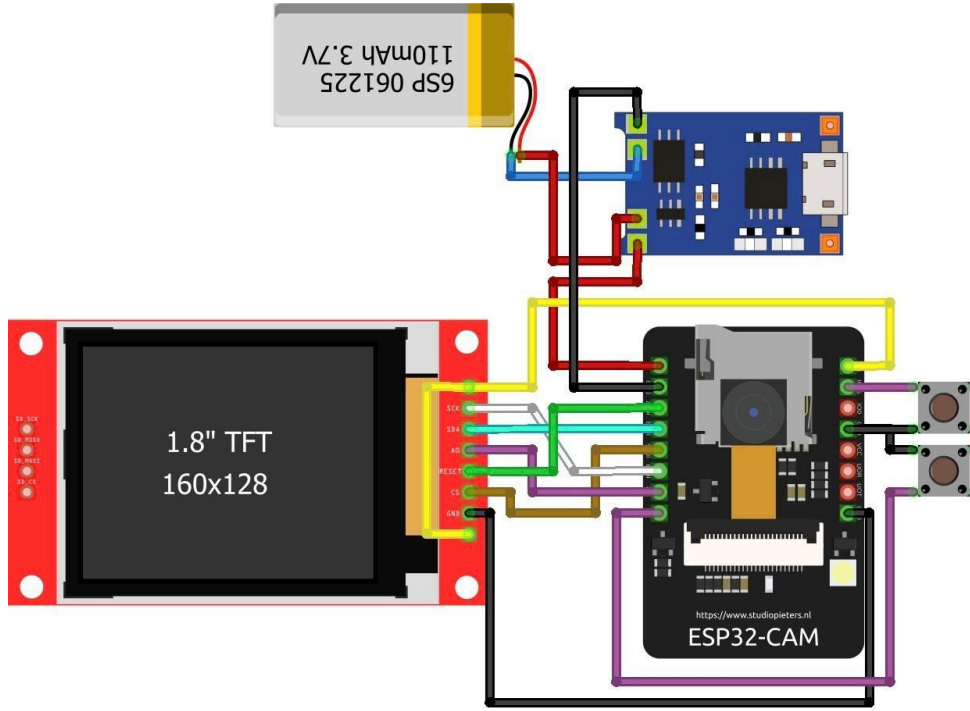
9.1.6 البنية التقنية لتطبيق NeoFaceMed



توضّح الصورة طريقة عمل تطبيق NeoFaceMed من الجانب التقني. فعند رفع صورة الوجه من الهاتف، تُرسل إلى خوادم Firebase حيث يتم تخزينها وتحليلها عبر نموذج الذكاء الاصطناعي، ثم تُعاد النتيجة إلى التطبيق ليعرضها للمستخدم بشكل فوري وواضح.

2.6 نموذج الجهاز التطبيقي

1.2.6 مخطط توصيل المكونات الإلكترونية



شكل 10.6: مخطط توصيل المكونات الإلكترونية

يبين هذا المخطط طريقة الربط بين المكونات الأساسية للجهاز. تُستخدم وحدة **ESP32-CAM** لالتقاط الصور ومعالجتها، بينما تُعرض النتائج على شاشة **TFT 8.1"**. تغذي بطارية الليثيوم **7V.3** الدائرة بالطاقة من خلال وحدة الشحن **TP4056** لضمان تشغيل مستقل وآمن للجهاز.

2.2.6 الواجهة الأمامية للجهاز



شكل 11.6: الجهاز من الأمام

يُظهر هذا الشكل الواجهة الأمامية للجهاز، والتي تحتوي على شاشة العرض لعرض النتائج مباشرة بعد المعالجة. تم تصميم هذه الواجهة بشكل بسيط وواضح لتسهيل الاستخدام، بينما توجد الكاميرا في الجهة الخلفية لالتقاط الصور أو الفيديو.

3.2.6 الواجهة الخلفية للجهاز



شكل 12.6: الجهاز من الخلف

يوضح هذا الشكل الجزء الخلفي للجهاز أثناء عملية التركيب، حيث تُثبت وحدة الكاميرا (ESP32-CAM) في الخلف، بينما تكون الشاشة في الواجهة الأمامية. كما يُظهر الشكل طريقة تثبيت المكونات الداخلية مثل البطارية ووحدة الشحن داخل الهيكل بشكل منظم وآمن لضمان

حماية المكونات وسهولة الصيانة.

المحور 7

نموذج العمل التجاري



شكل 1.7: نموذج العمل التجاري

خاتمة

في ختام هذا المشروع، يمكن القول إنه يمثل إسهامًا حقيقيًا في مجال توظيف الذكاء الاصطناعي لخدمة الصحة العامة، من خلال الجمع بين التقنيات الحديثة والرؤية الإنسانية الهادفة إلى تسهيل الكشف المبكر عن الحالات الصحية ودعم عمل الطاقم الطبي بوسائل رقمية ذكية وسهلة الاستخدام. لقد سعينا من خلال هذا الابتكار إلى إثبات أن التحول الرقمي في القطاع الصحي لا يقتصر على الأنظمة المعقدة أو الأجهزة الضخمة، بل يمكن تحقيق أثر ملموس عبر حلول بسيطة، دقيقة، وقابلة للتطبيق في مختلف البيئات الصحية.

إن المشروع لا يهدف فقط إلى عرض نموذج تقني أو فكرة برمجية، بل يتجاوز ذلك نحو إرساء مفهوم الرعاية الوقائية الذكية التي تعتمد على البيانات والتحليل العلمي بدلاً من الملاحظة التقليدية. فهو مشروع يسعى لأن يكون صلة وصل بين العلم والمجتمع، وبين الذكاء الاصطناعي والإنسان، ليجعل من التكنولوجيا وسيلة لخدمة الحياة لا غاية في حد ذاتها.

ورغم الجهود الكبيرة التي بُذلت في جميع المراحل، فقد واجهنا تحديات متعددة تمثلت في محدودية الموارد التقنية وصعوبة الوصول إلى مكونات إلكترونية مناسبة، إضافة إلى غياب قواعد بيانات طبية دقيقة يمكن الاعتماد عليها في تدريب النماذج الذكية وتحسين دقتها. إلا أن هذه العقبات لم تُثنيّا، بل شكّلت دافعًا قويًا نحو البحث والتعلم والتطوير، فكلّ تحدٍّ واجهناه كان خطوة إضافية نحو بناء رؤية أكثر وضوحًا ونضجًا.

نحن نؤمن بأنّ هذا المشروع لا يمثل نهاية المطاف، بل بداية مسار طويل نحو منظومة صحية رقمية أكثر كفاءة وعدلاً، تعتمد على الذكاء الاصطناعي في التشخيص المبكر، وتضع الأسس لثقافة طبية رقمية قائمة على التنبؤ والوقاية بدل العلاج اللاحق. كما نطمح مستقبلاً إلى توسيع نطاق العمل عبر تطوير النموذج الأولي وتحسين قدراته التحليلية، وإرساء شراكات بحثية مع مؤسسات ومخابر متخصصة، ليصبح المشروع منتجًا وطنيًا رائدًا يمكن تعميمه على نطاق أوسع.

في النهاية، يظل هذا العمل ثمرة جهد علمي وإنساني متكامل، جمع بين الشغف بالمعرفة والرغبة الصادقة في الإسهام في رفاه المجتمع. وإنّ نجاحه الحقيقي لا يُقاس بما أنجزناه فحسب، بل بما فتحه من آفاق جديدة للتفكير، والابتكار، والإيمان بأنّ العلم حين يُسخر لخدمة الإنسان، يصنع مستقبلاً أكثر رحمة وذكاءً وعدالة.