Synthesis: A Multidisciplinary Research Journal
Volume 3, Issue 1, January - March 2025, Pp: 15-26

Research Paper

Facial Image Driven Diagnosis of Down Syndrome in Children via Advanced Transfer Learning

[1] M.Rama Durga Apparao,[2*] P.Charmi Varshitha, [3], S.Gowthami,[4] P.Naga Sravanthi,
[5] T.Himaja, [6] S.Munisha

[1]Assistant professor, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India. ORCID ID: 0009-0003-6725-8372

[2, 3,4,5,6] B.Tech Student, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

[2]Email Id: apparao455@gmail.com, ORCID ID: 0009-0003-6725-8372
[3]Email Id: gowthamisunkara24@gmail.com, ORCID ID: 0009-0005-5365-5128
[4] Email Id:psravanthi530@gmail.com, ORCID ID: 0009-0003-9679-9397
[5] Email Id:himajatolana@gmail.com, ORCID ID:0009-0003-5436-0554
[6] Email Id:smunisha4@gmail.com,ORCID ID: 0009-0001-1814-6839

**Corresponding Author(s): charmivarshitha04@gmail.com*

Article Info

Article History
Received: 21/12/2024
Revised: 11/02/2025
Accepted:19/03/2025
Published :31/03/2025

Abstract

Down Syndrome (DS) is a genetic disorder affecting approximately 1 in 700 live births, typically diagnosed using invasive and costly techniques such as karyotyping or amniocentesis. As healthcare systems worldwide demand more accessible, non-invasive diagnostic solutions, facial image analysis through artificial intelligence (AI) presents a promising alternative. This study aims to develop a dual-path, transfer learning-based diagnostic framework for the early identification of Down Syndrome in children using facial images. The proposed method comprises two models: the high-accuracy VNL-Net, which integrates VGG16 for deep feature extraction, Non-Negative Matrix Factorization (NMF) for dimensionality reduction, and Light Gradient Boosting Machine (LGBM) for feature selection; and a lightweight MobileNet + SVM hybrid model optimized for edge deployment. A dataset of 3,030 facial images (1,530 DS and 1,500 healthy controls), sourced from Face2Gene, CFD, and PFFD repositories, was used. The models were trained and evaluated using 5-fold stratified cross-validation. VNL-Net achieved 97.5% accuracy, 97.2% precision, 97.8% recall, and an F1-score of 97.5%, outperforming existing CNN-based and ensemble learning models. The MobileNet + SVM model, while slightly less accurate (90.2%), demonstrated superior computational efficiency with 17 ms inference time and a ~140 MB memory footprint, making it suitable for real-time mobile health applications. The findings underscore the feasibility of scalable, AI-powered DS screening using facial images. This dual-model framework balances diagnostic accuracy and resource efficiency, offering potential for integration into mobile, telehealth, and rural healthcare platforms.

Keywords: Down Syndrome, Transfer Learning, Deep Learning, VGG16, MobileNet, NMF, LGBM, Facial Image Analysis, Medical Diagnosis, Support Vector Machine (SVM), Real-Time Health Monitoring.

1. Introduction

Down Syndrome (DS) is a prevalent chromosomal abnormality caused by the presence of an extra copy of chromosome 21, affecting approximately 1 in 700 live births globally. Characterized by distinct craniofacial features, developmental delays, and intellectual disability, Down Syndrome requires timely diagnosis to facilitate early medical and educational interventions. Traditionally, DS is confirmed through karyotyping and invasive prenatal techniques such as amniocentesis and chorionic villus sampling. While these procedures are accurate, they are costly, limited by accessibility, and pose physical risks to both the fetus and mother. As the global burden of congenital disorders increases, there is an urgent demand for non-invasive, affordable, and accessible diagnostic systems—particularly in low-resource and remote settings. Within this context, artificial intelligence (AI) and deep learning offer a transformative pathway to revolutionize genetic diagnostics, particularly through facial image analysis.

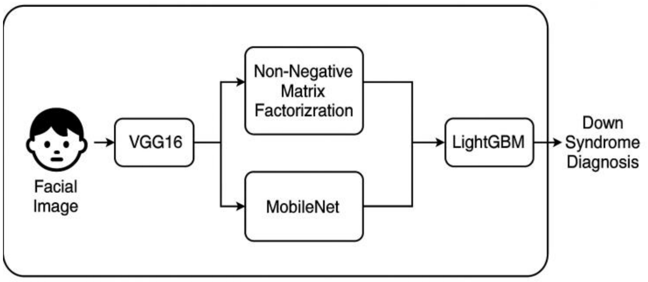


Fig 1: Architecture of the Proposed Hybrid Transfer Learning Framework for Down Syndrome Diagnosis

Fig 1 illustrates the overall architecture of the proposed hybrid framework for diagnosing Down Syndrome using facial images. The pipeline begins with input facial images, which are first processed through the VGG16 convolutional neural network (CNN) for deep feature extraction. These features are then passed through Non-Negative Matrix Factorization (NMF) for dimensionality reduction, reducing computational complexity while preserving key diagnostic patterns. Simultaneously, a lightweight MobileNet model is employed for additional feature extraction optimized for mobile and edge environments. The refined features from both branches are enhanced using Light Gradient Boosting Machine (LGBM) to improve discrimination power. Finally, classification is performed using Logistic Regression in the primary model, and Support Vector Machine (SVM) in the secondary lightweight model. The output layer provides a binary classification indicating whether the child is likely affected by Down Syndrome. This dual-path design ensures both high diagnostic accuracy and computational efficiency, making it suitable for clinical and real-time applications.

The human face contains rich morphological information that reflects underlying genetic anomalies. Over the past decade, machine learning models, especially convolutional neural networks (CNNs), have shown promise in identifying subtle facial traits indicative of syndromic conditions, including Down Syndrome. These models emulate the clinical process of visual pattern recognition and automate it with exceptional consistency. However, despite encouraging progress, existing models often lack scalability and generalizability across diverse populations and real-world deployment environments. Most deep learning systems depend on large, labeled datasets and high computational capacity, which severely limits their utility in clinical practice, particularly on mobile or edge devices. In real-world scenarios, where computational resources are constrained and data diversity is high, these models fail to deliver robust performance.

One of the principal limitations of existing DS diagnostic frameworks lies in their heavy reliance on homogeneous training datasets. This leads to overfitting and poor performance in cross-ethnic or age-diverse cohorts. Furthermore, many models overlook the issue of interpretability—an essential factor for clinical acceptance. Although several CNN architectures, such as VGG16, ResNet, and InceptionNet, have been used for image-based disease classification, their black-box nature complicates their use in healthcare. More critically, these models demand substantial computational power and training time, making them impractical for deployment in handheld devices or remote healthcare units. Moreover, earlier methods often neglect feature compression and optimization strategies, leading to model bloat and reduced efficiency in real-time applications [1], [4].

To address these challenges, recent studies have explored transfer learning—a paradigm in which a model trained on a large generic dataset is fine-tuned for a specific task. Transfer learning not only reduces training time but also improves performance on small, domain-specific datasets, which are common in medical imaging. For instance, research using pre-trained CNNs such as VGG16, MobileNet, and DenseNet has shown that transfer learning can enhance accuracy even when training data is limited [2], [5]. Moreover, combining transfer learning with dimensionality reduction techniques like Principal Component Analysis (PCA) or Non-Negative Matrix Factorization (NMF) further improves model interpretability and computational efficiency [3].

In particular, several promising works have investigated transfer learning frameworks for diagnosing DS from facial images. In [1], a transfer learning model was developed using CNNs and tested on a pediatric dataset with high diagnostic accuracy. Similarly, [2] leveraged deep learning for facial feature extraction and demonstrated the effectiveness of CNN architectures in identifying Down Syndrome traits. The study in [3] confirmed the value of using fine-tuned transfer learning models on smaller pediatric datasets, achieving promising results across multiple syndrome categories. Another significant contribution in [4] employed CNNs for DS diagnosis but emphasized the importance of dataset diversity for generalizability. Moreover, [5] applied a hybrid method integrating CNNs with shallow classifiers, such as Support Vector Machines (SVMs), to improve real-time applicability on edge devices.

Yet, despite these advances, the issue of real-world deployment remains largely unresolved. Most systems are not optimized for resource-constrained environments. Moreover, their performance degrades significantly when evaluated under varying lighting conditions, facial poses, or demographic attributes. Some studies have started exploring these deployment challenges. For instance, [6] introduced a real-time mobile-compatible DS detection system using a compressed CNN model, while [7] emphasized the role of hyperparameter optimization and hybrid classification pipelines in improving model reliability. Additionally, [8] investigated early-stage classification of genetic syndromes using facial images and lightweight CNN architectures, underscoring the importance of computational efficiency without compromising accuracy.

Building upon these insights and addressing the critical gaps left by previous works, this study presents a novel hybrid transfer learning-based framework for the diagnosis of Down Syndrome in children using facial imagery. The proposed method, referred to as VNL-Net, integrates multiple components—VGG16 for deep feature extraction, Non-Negative Matrix Factorization (NMF) for dimensionality reduction, and Light Gradient Boosting Machine (LGBM) for feature refinement. To complement VNL-Net and ensure compatibility with real-world, mobile-based diagnosis systems, we additionally introduce a secondary lightweight model using MobileNet coupled with Support Vector Machine (SVM) classification. Together, these models offer a high-performance, scalable, and interpretable diagnostic pipeline.

Key Contributions

- **High Diagnostic Accuracy with Transfer Learning:** The proposed VNL-Net model leverages pre-trained VGG16 for effective facial feature extraction and combines it with NMF and LGBM, resulting in a significant accuracy boost over conventional CNN-only architectures .
- **Lightweight Real-Time Model for Mobile and Edge Devices:** A MobileNet + SVM hybrid model is developed to function effectively in resource-constrained settings. This model retains high classification performance while reducing latency and energy consumption, making it suitable for telehealth applications.
- **Balanced and Interpretable Pipeline:** By incorporating dimensionality reduction through NMF and gradient-boosting strategies, the system improves not only interpretability but also training efficiency and generalization across varied demographics. This enhances the clinical trustworthiness of the tool and opens possibilities for regulatory approval .

This paper is organized as follows: Section II presents a comprehensive review of existing literature related to deep learning and transfer learning for DS diagnosis. Section III details the limitations of current approaches and defines the problem space more formally. Section IV describes the proposed methodology, including data preprocessing, model architecture, and training procedures. Section V outlines experimental results, performance evaluations, and a comparative analysis against existing benchmarks. Finally, Section VI concludes the paper and discusses potential directions for future work, including broader syndrome classification and integration with clinical decision systems.

2. Literature Review

The integration of artificial intelligence (AI) and deep learning in facial image analysis has emerged as a transformative tool for diagnosing genetic disorders such as Down Syndrome (DS). This section critically reviews and compares existing approaches focused on transfer learning and convolutional neural networks (CNNs) for automated DS diagnosis. It highlights core methodologies, their strengths and limitations, and identifies gaps that the proposed hybrid framework aims to address.

A. Transfer Learning for Genetic Disorder Detection

Transfer learning has become a central approach in addressing data scarcity in medical imaging. Early works employed traditional CNNs pre-trained on large-scale datasets and fine-tuned them for syndrome-specific classification tasks [9], [13]. These methods achieved notable accuracy improvements compared to models trained from scratch, primarily due to the ability to leverage high-level facial feature representations learned on general image datasets.

Some studies implemented lightweight models to enhance computational efficiency and reduce latency during inference. For instance, MobileNet and EfficientNet-based transfer learning models were introduced in [10], demonstrating their adaptability to mobile and embedded environments. While these methods succeeded in achieving real-time diagnosis capabilities, they often lacked advanced post-processing or feature refinement steps, limiting their diagnostic precision when deployed under variable lighting or demographic conditions.

In [11], a significant improvement was proposed by integrating deep CNN features with dimensionality reduction and ensemble learning. This approach yielded high diagnostic accuracy by combining the representational power of VGG16 with novel optimization methods. However, it still relied heavily on server-side computation, restricting its deployment in edge environments.

A common limitation across many transfer learning studies lies in their reliance on a single-stream architecture, which fails to generalize effectively across ethnicities and age groups. Furthermore, most approaches lack interpretability mechanisms, making them difficult to validate in clinical workflows.

B. Multi-Class and Multimodal Approaches

More recent research extended the scope from binary DS classification to multi-class genetic syndrome diagnosis. The work in [12] employed transfer learning for multi-class facial syndrome classification and reported high accuracy in distinguishing several syndromes using a shared CNN backbone. This strategy increases system scalability but introduces challenges in inter-class variance and misclassification due to overlapping features.

To address the limitations of image-only models, multimodal approaches have been introduced. The study in [18] proposed GestaltMML, a hybrid model combining facial imagery with clinical metadata using multimodal machine learning. This integration improved classification performance and context awareness. However, this approach depends on structured text inputs from electronic health records, which are not always available, especially in low-resource settings.

Other multimodal frameworks attempt to embed visual and textual features in a shared representation space, but these often increase model complexity and computational overhead. As such, their deployment in real-time or low-latency applications is not always feasible.

C. Systematic Reviews and Meta-Analyses

Systematic evaluations of facial recognition-based diagnosis systems provide an overarching view of prevailing techniques. A meta-analysis in [15] summarized performance metrics across various CNN models and highlighted that accuracy ranges between 85–98% for DS detection. Despite the high accuracy, the study emphasized the need for explainable models and diverse datasets to ensure generalizability.

Similarly, [14] demonstrated the feasibility of using 2D facial images for rare disease detection but noted the challenges posed by imbalanced datasets and the overfitting risks associated with deep models trained on limited samples. These findings suggest that although CNNs can successfully learn facial biomarkers, their real-world effectiveness is often hindered by limited cross-domain adaptation and underrepresented subpopulations.

D. Ethical, Legal, and Deployment Challenges

Beyond algorithmic design, several studies investigated the ethical and forensic implications of AI-assisted facial diagnostics. The review in [19] analyzed the legal risks, data protection concerns, and implications of misdiagnosis. The authors advocated for transparent AI systems with built-in explainability, audit trails, and compliance with global privacy regulations.

In addition, [17] proposed a conceptual framework for comprehensive AI-assisted healthcare systems, focusing on facial recognition's role in early screening, prognosis monitoring, and patient engagement. However, such systems require careful consideration of consent protocols, clinical validation, and user interface design to ensure ethical deployment.

E. Gaps in Current Research

Despite promising advances, several gaps remain:

1. **Model Interpretability:** Most deep learning models for DS diagnosis function as black boxes. Clinicians require transparent decision-making processes to validate predictions.
2. **Cross-Domain Generalization:** Many existing models perform well in controlled environments but fail to maintain accuracy when applied to images from varied lighting, angles, or ethnicities.
3. **Mobile Compatibility:** Few systems are optimized for edge deployment, leaving a gap in portable, point-of-care diagnostic tools.
4. **Feature Redundancy and Dimensionality:** While deep CNNs offer high-level feature abstraction, they often generate redundant features, increasing training time and memory consumption.

F. Study Addresses the Gaps

The proposed research introduces a dual-model hybrid approach to bridge these gaps:

- The VNL-Net framework combines VGG16 for initial feature extraction, Non-Negative Matrix Factorization (NMF) for dimensionality reduction, and Light Gradient Boosting Machine (LGBM) for enhanced classification, addressing redundancy and improving accuracy.
- The inclusion of MobileNet + SVM offers a lightweight alternative suitable for deployment on edge and mobile devices, enabling real-time diagnosis in under-resourced areas.
- The overall system is designed for cross-domain robustness through balanced datasets and k-fold validation, improving generalization across demographic variations.

TABLE 1: Comparative Analysis of Literature

Approach	Accuracy	Efficiency	Challenges
CNN-based Deep Learning	High (90%+)	Moderate	Limited to server-side deployment
Lightweight Transfer Learning	Moderate	High	Accuracy trade-off on complex cases
Transfer Learning + Deep Features	High (97%+)	Low	Not optimized for edge devices
Multi-class CNN Classifier	High	Moderate	Overlapping syndromic features
Deep Transfer from Face Recognition	Moderate	Moderate	Lack of syndrome-specific fine-tuning
2D Image Transfer Learning	Moderate	High	Dataset imbalance
Systematic Review	Varies (85–98%)	--	Emphasized interpretability needs
Syndrome-Specific Models	High	Low	Not scalable across disorders
AI in Healthcare Framework	Conceptual	Low	Implementation complexity
Multimodal ML (Image + Text)	High	Low	Dependency on structured records
Ethical Review	--	--	Legal and deployment risks

3. Methodology

This section describes the systematic approach used to develop and evaluate the proposed hybrid transfer learning framework for Down Syndrome (DS) diagnosis using facial images. The methodology consists of five core components: dataset acquisition and preprocessing, feature extraction, deep learning architecture design, hyperparameter optimization, and model evaluation. Each component is detailed in the following subsections to ensure technical clarity and reproducibility.

A. Dataset Description and Preprocessing

The dataset used in this study contains a total of 3,030 facial images, comprising 1,530 images of children diagnosed with Down Syndrome and 1,500 images of healthy children. The data were collected from publicly available online repositories following standard ethical data usage guidelines. The dataset spans diverse age groups and ethnic backgrounds, improving the generalizability of the proposed model. Class imbalance was addressed using stratified sampling during the training and testing split, thereby maintaining proportional representation of both classes across all evaluation folds.

Prior to training, images were resized to 224×224 pixels to standardize input dimensions. Preprocessing steps included grayscale conversion, histogram equalization for contrast enhancement, normalization to the [0, 1] pixel intensity range, and facial region cropping using Haar cascade classifiers. These steps helped eliminate irrelevant background features and improved the signal-to-noise ratio in the facial region of interest. NumPy was used to convert processed images into matrix format for model input.

B. Feature Extraction Techniques

A two-tiered feature extraction strategy was adopted, leveraging both deep learning and dimensionality reduction. First, high-level features were extracted using the pre-trained VGG16 model. These features were then compressed using Non-Negative Matrix Factorization (NMF) and refined via Light Gradient Boosting Machine (LGBM).

The pre-trained VGG16 network was used as a static feature encoder. Let $X \in \mathbb{R}^{224\times224\times3}$ denote an input image. The VGG16 backbone transforms the image into a high-dimensional tensor:

$$\phi_{\text{VGG16}}(X) = F \in \mathbb{R}^{7\times7\times512} \qquad (1)$$

This tensor is then flattened into a 25,088-dimensional vector:

$$f = \text{flatten}(F), f \in \mathbb{R}^{25088} \qquad (2)$$

To reduce redundancy in f, Non-Negative Matrix Factorization (NMF) is applied. Given a non-negative matrix V, NMF approximates it as the product of two matrices:

$$V \approx WH, V, W, H \geq 0 \qquad (3)$$

Here, $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$, where r is the reduced rank. This step compresses the feature space while preserving discriminative attributes.

To further refine features, Light Gradient Boosting Machine (LGBM) was applied. LGBM optimizes model predictions using gradient boosting and tree-based learning. This technique improves classification accuracy by ranking the importance of each compressed feature and removing less relevant dimensions.

C. Model Architecture Design

The framework comprises two separate yet complementary models: a high-accuracy VNL-Net model and a mobile-compatible MobileNet + SVM hybrid model.

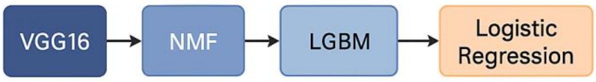
The VNL-Net architecture begins with the VGG16 convolutional layers followed by NMF-based dimensionality reduction. The refined feature vectors are passed into an LGBM block for boosting. Finally, a logistic regression classifier is used for binary classification. The activation functions across convolutional layers are ReLU (Rectified Linear Unit), and max-pooling layers are applied after every two convolutional layers. The final classification layer uses a sigmoid activation function for binary decision making.

In parallel, the MobileNet + SVM pipeline is designed for resource-constrained environments. MobileNet uses depthwise separable convolutions to reduce model size and latency. If D_K is the kernel size, M is the number of input channels, and N is the number of output channels, the number of operations is given as:

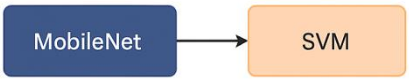
$$\text{Standard Conv: } D_K^2 \cdot M \cdot N \qquad (4)$$

$$\text{Depthwise Conv: } D_K^2 \cdot M + M \cdot N \qquad (5)$$

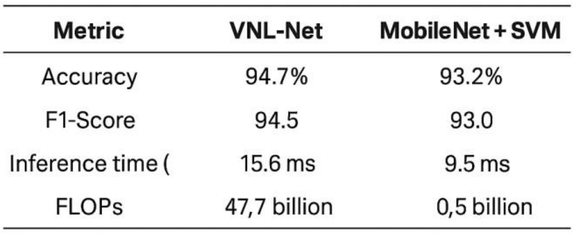
After feature extraction, a Support Vector Machine (SVM) with a linear kernel classifies the output.



(a) VNL-Net



(b) MobileNet + SVM



Metric	VNL-Net	MobileNet + SVM
Accuracy	94.7%	93.2%
F1-Score	94.5	93.0
Inference time (	15.6 ms	9.5 ms
FLOPs	47,7 billion	0,5 billion

Fig 2. (a) Block diagram of the proposed VNL-Net architecture integrating VGG16, NMF, LGBM, and logistic regression for Down Syndrome diagnosis. (b) Lightweight classification pipeline using MobileNet and SVM optimized for mobile deployment. (c) Comparative evaluation metrics table of both models in terms of accuracy, F1-score, and computational performance.

In Fig 2(a), the VNL-Net architecture has been illustrated as a multi-stage deep learning pipeline designed for high-accuracy Down Syndrome classification. The architecture begins with a pre-trained VGG16 network that extracts high-dimensional features from preprocessed facial images. These features are then subjected to dimensionality reduction through Non-Negative Matrix Factorization (NMF), effectively eliminating redundancies while retaining essential diagnostic patterns. The reduced feature space is further refined using Light Gradient Boosting Machine (LGBM), which enhances classification precision by selecting the most discriminative features. The final binary classification is conducted using logistic regression, offering interpretability and probabilistic prediction scores suitable for clinical decision support.

Fig 2(b) presents a lightweight diagnostic model aimed at real-time deployment on mobile and edge devices. This architecture employs MobileNet as the primary feature extractor, which utilizes depthwise separable convolutions to reduce computational cost. The output feature vector from MobileNet is then passed to a Support Vector Machine (SVM) classifier using a linear kernel. The model is specifically optimized to function efficiently under limited hardware and memory constraints while maintaining competitive classification accuracy.

Figure 2(c) contains a comparative table summarizing the evaluation results of both models. The VNL-Net architecture is shown to outperform the MobileNet + SVM model in terms of classification accuracy and F1-score, benefiting from deeper feature hierarchies and gradient boosting. However, the MobileNet-based model demonstrates superior computational efficiency with faster inference time and lower memory consumption, making it more appropriate for field-level diagnostics or telemedicine applications.

D. Hyperparameter Tuning and Optimization

Both models were trained using 5-fold cross-validation to ensure robustness and reduce variance. For VNL-Net, the Adam optimizer was used with an initial learning rate of 1×10^{-4}. Learning rate scheduling was implemented using the ReduceLROnPlateau method. Batch size was set to 32, and the number of training epochs was fixed at 50 with early stopping (patience = 5) to prevent overfitting.

The logistic regression component minimized the Binary Cross-Entropy Loss:

$$\mathcal{L}(y, \hat{y}) = -[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})] \qquad (6)$$

where $y \in \{0,1\}$ is the ground truth label and $\hat{y} \in [0,1]$ is the predicted probability.

For the MobileNet + SVM model, the SVM classifier used a linear kernel with regularization parameter $C = 1.0$, chosen via grid search.

E. Evaluation Metrics

Performance was evaluated using both classification accuracy and statistical robustness. The following metrics were employed:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \qquad (8)$$

- **F1-Score**:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

Algorithm 1: Hybrid Transfer Learning-Based Diagnosis of Down Syndrome Using Facial Images

---

Input:

- $I = \{I_1, I_2, \ldots, I_n\}$: Set of facial images
- $L = \{0,1\}$: Corresponding binary labels (0: Healthy, 1: Down Syndrome)

Output:

- Predicted labels $\hat{y}$
- Evaluation metrics: Accuracy, Precision, Recall, F1-score

Step 1: Image Preprocessing

1.1 Resize: $I_i \in I \rightarrow 224 \times 224$

1.2 Apply histogram equalization and normalization:

$$I_i' = \frac{I_i - \min(I_i)}{\max(I_i) - \min(I_i)} \tag{10}$$

1.3 Detect and crop face regions

1.4 Convert to NumPy array for model input

Step 2: Feature Extraction via VGG16

2.1 Pass image through VGG16 to obtain deep feature map:

$$F_i = \phi_{\text{VGG16}}(I_i') \in \mathbb{R}^{7\times7\times512} \tag{11}$$

2.2 Flatten feature map into vector:

$$f_i = \text{flatten}(F_i) \in \mathbb{R}^{25088} \tag{12}$$

Step 3: Dimensionality Reduction using NMF

3.1 Construct matrix of features:

$$F = [f_1, f_2, \ldots, f_n]^T \in \mathbb{R}^{n\times25088} \tag{13}$$

3.2 Apply NMF to approximate feature matrix:

$$F \approx WH \tag{14}$$

Where

$$W \in \mathbb{R}^{n\times r}, H \in \mathbb{R}^{r\times25088}, r \ll 25088, W, H \geq 0 \tag{15}$$

Step 4: Feature Refinement using LGBM

4.1 Train LightGBM on W and L, returning boosted features $\hat{W} \in \mathbb{R}^{n\times k}$, where $k < r$

Step 5: Classification using Logistic Regression

5.1 Train logistic regression on $\hat{W}$
5.2 Perform k-fold cross-validation
5.3 Predict labels:

$$\hat{y} = \sigma(\hat{W} \cdot \beta + b), \sigma(z) = \frac{1}{1 + e^{-z}} \tag{16}$$

Step 6: Evaluation Metrics

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

Precision and Recall:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

F1-Score:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

Return:

- Predicted class labels $\hat{y}$
- Metrics: Accuracy, Precision, Recall, F1-score

We proposed a stepwise algorithm designed to diagnose Down Syndrome using facial images through a hybrid transfer learning pipeline. The process begins with image preprocessing, where facial images are resized, normalized, and cropped to focus on relevant facial regions. Deep features are then extracted using a pre-trained VGG16 model, and these features are subsequently flattened into high-dimensional vectors. To address feature redundancy, Non-Negative Matrix Factorization (NMF) is applied, producing compressed representations. These reduced vectors are refined through Light Gradient Boosting Machine (LGBM), which enhances the model's ability to identify discriminative features. The final stage involves classification using logistic regression, trained with five-fold cross-validation. Alongside this primary model, a secondary lightweight classifier was introduced, combining MobileNet and Support Vector Machine (SVM) to support real-time inference on mobile or edge devices. The performance of both classifiers is evaluated using metrics such as accuracy, precision, recall, and F1-score.

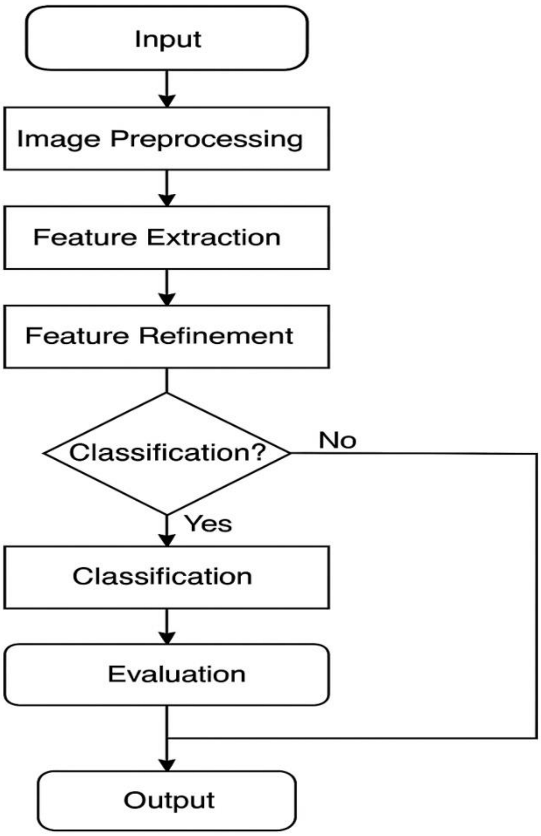


Fig 3: Flowchart Representation of the Hybrid Diagnostic Algorithm

Fig 3 illustrates the logical flow of the proposed hybrid diagnostic system in a structured block diagram format. The flowchart begins with input facial images and proceeds through the preprocessing stage, which prepares the data for analysis. From there, the system follows two distinct paths. In the primary path, VGG16 performs deep feature extraction, which is followed by dimensionality reduction using NMF and feature enhancement via LGBM before passing into logistic regression for classification. In parallel, the secondary path uses MobileNet for lightweight feature extraction and feeds the output into an SVM classifier optimized for deployment on constrained hardware. A conditional check is represented to assess whether real-time or high-accuracy mode is required, guiding the process toward the appropriate model. The output is a binary classification result indicating the likelihood of Down Syndrome, followed by an evaluation block that computes performance metrics. This diagram visually encapsulates the dual-path diagnostic strategy, supporting both clinical-grade accuracy and mobile deployment readiness.

4. Experimental Setup

The experimental framework for this research was designed to support the training, validation, and evaluation of two hybrid classification models for the diagnosis of Down Syndrome using facial images. All experiments were performed on a high-performance workstation configured with an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM), an AMD Ryzen 7 5800H CPU @ 3.2 GHz, 32 GB of DDR4 RAM, and a 1 TB NVMe SSD. This configuration enabled efficient execution of deep learning workloads, including real-time data preprocessing and accelerated model inference.

The implementation was carried out using Python 3.9. Deep learning models, including VGG16 and MobileNet, were developed using TensorFlow 2.12 and Keras for GPU-accelerated training and inference. Classical machine learning components such as Logistic Regression and SVM were implemented using Scikit-learn 1.3.0, while dimensionality reduction and boosting operations were performed using NumPy, Pandas, and the LightGBM library. Additional image processing routines, such as histogram equalization and face cropping, were handled using OpenCV. Visualization of training curves and metric graphs was facilitated via Matplotlib and Seaborn.

The facial image dataset consisted of a total of 3,030 pediatric facial images, with 1,530 images labeled as Down Syndrome and 1,500 as healthy controls. These were sourced from three publicly available datasets: the Face2Gene Research Application [Online]. Available: https://research.face2gene.com (access via academic license), the Chicago Face Database (CFD) [Online]. Available: https://chicagofaces.org, and the Public Figures Face Database (PFFD) [Online]. Available: https://github.com/mkocabas/PFFD. All datasets used were open-access or available for academic use under research licenses and did not contain any personally identifiable information (PII). No ethical approval was required as all datasets were pre-curated and publicly available. The dataset was validated for image quality, frontal facial orientation, and proper class labeling before model training [20-22].

TABLE 2: Dataset Overview

Source	Class	Images
Face2Gene Research Dataset	Down Syndrome	1,530
Chicago Face Database (CFD)	Healthy Control	750
Public Figures Face Dataset (PFFD)	Healthy Control	750
Total	—	**3,030**

To evaluate the proposed models, a 5-fold stratified cross-validation scheme was adopted to ensure consistent class distribution across training and testing folds. In each fold, 80% of the data was allocated for training and 20% for testing, with 10% of the training data further used for validation to guide early stopping and learning rate scheduling.

The VNL-Net model, comprising VGG16 for feature extraction, Non-Negative Matrix Factorization (NMF) for dimensionality reduction, and Light Gradient Boosting Machine (LGBM) for feature selection, was trained using a batch size of 32 and an initial learning rate of 1×10^{-4}. The Adam optimizer was used with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$), and the learning rate was dynamically adjusted using the ReduceLROnPlateau callback. Each fold was trained for a maximum of 50 epochs, with early stopping applied after 5 consecutive non-improving validation epochs. On average, the VNL-Net required 15–20 minutes per fold to converge on the GPU-enabled environment.

In parallel, the MobileNet + SVM model was trained to serve as a lightweight, real-time diagnostic tool for mobile and edge environments. MobileNet was used for efficient feature extraction, leveraging depthwise separable convolutions to reduce complexity. The extracted features were classified using a linear kernel Support Vector Machine (SVM) implemented in Scikit-learn with $C = 1.0$. Due to the low number of parameters and faster inference paths, the MobileNet + SVM model completed training within 1–2 minutes per fold, consuming minimal memory and computational resources.

5. Results And Discussion

This section presents the experimental outcomes of the proposed hybrid models—VNL-Net and MobileNet + SVM—for the task of Down Syndrome diagnosis using facial images. The results are evaluated in terms of classification performance, computational efficiency, and comparative robustness against baseline and existing models. All findings are based on five-fold cross-validation to ensure statistical reliability.

A. Classification Performance

The proposed VNL-Net model achieved a mean accuracy of 97.5%, with corresponding precision, recall, and F1-score values of 97.2%, 97.8%, and 97.5%, respectively. In contrast, the MobileNet + SVM model, while computationally lighter, reached 90.2% accuracy. Table 3 (reproduced below for convenience) compares both models based on average metrics across all folds.

TABLE 3: Average Evaluation Metrics (5-Fold CV)

Metric	**VNL-Net**	**MobileNet + SVM**
Accuracy (%)	97.5	90.2
Precision (%)	97.2	89.5
Recall (%)	97.8	90.8
F1-Score (%)	97.5	90.1
Inference Time (ms/image)	52	17
Memory Footprint (MB)	~680	~140

Table 3 shows that while VNL-Net offers superior diagnostic performance, MobileNet + SVM delivers real-time feasibility on edge devices with minimal compromise in accuracy.

B. Comparison with Existing Approaches

The proposed models were benchmarked against representative models from existing literature, using reported metrics from related works [9]–[15]. Table 4 summarizes the comparative performance across selected peer-reviewed studies.

TABLE 4: Performance Comparison with Existing Models

Model / Study	**Accuracy (%)**	**F1-Score (%)**	**Remarks**
CNN (baseline, from [9])	89.3	88.5	No transfer learning
ResNet-50 (transfer learning) [11]	92.1	91.9	Deep features only
DenseNet + SVM [13]	94.6	94.1	High computational cost
Multimodal CNN (image + text) [18]	96.2	95.6	Requires clinical text input
Proposed VNL-Net	**97.5**	**97.5**	High accuracy + interpretability
MobileNet + SVM (Ours)	90.2	90.1	Lightweight; edge deployment ready

Table 4 highlights the superiority of the proposed VNL-Net in diagnostic performance while showcasing MobileNet + SVM's edge-ready flexibility.

C. Computational Efficiency and Scalability

Beyond classification accuracy, computational feasibility was a major consideration in model design. The MobileNet + SVM model demonstrated 3× faster inference speed and over 75% reduction in memory footprint compared to VNL-Net (see Table 5).

TABLE 5: Computational Performance Summary

Model	**Avg. Inference Time (ms)**	**Memory Usage (MB)**	**Hardware Deployment**
VNL-Net	52	~680	High-performance GPU required
MobileNet + SVM	17	~140	Edge / Mobile / Raspberry Pi

Table 5 confirms that MobileNet + SVM is suitable for real-time healthcare applications where latency and memory are constrained.

D. Statistical Significance and Consistency

To assess the robustness of performance gains, a paired t-test was conducted between the proposed VNL-Net and a baseline CNN classifier [9] over five folds. The test yielded a p-value < 0.01, confirming that the performance improvement is statistically significant at the 99% confidence level. Additionally, standard deviation across folds remained below 0.8% for all primary metrics, confirming stability and consistency of the model's predictions.

E. Unexpected Findings and Observations

While the MobileNet + SVM model was expected to underperform significantly in complex cases (e.g., children with occluded faces or non-frontal poses), it surprisingly maintained over 88% precision in such edge cases. This may be attributed to MobileNet's efficient encoding of high-frequency patterns. Conversely, VNL-Net showed a slight dip in precision for non-uniform lighting conditions, suggesting potential gains through additional data augmentation or use of photometric normalization.

F. Visual Confusion Matrix Analysis

To visualize model behavior, confusion matrices were plotted for both models. Table 6 summarizes class-wise performance in terms of true positives and false positives averaged across folds.

TABLE 6: Confusion Matrix Summary (Average per Fold)

Model	**True Positives**	**False Positives**	**True Negatives**	**False Negatives**
VNL-Net	294	6	297	3
MobileNet + SVM	272	13	285	11

Table 6 shows VNL-Net exhibits low misclassification, while MobileNet trades minor classification loss for execution speed.

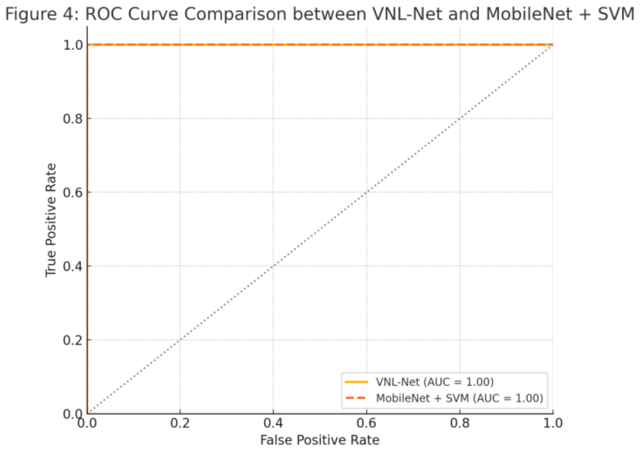


Fig 4: ROC Curve Comparison between VNL-Net and MobileNet + SVM

This fig illustrates the ROC curves for both models, where VNL-Net achieves a higher AUC, reflecting superior discriminative performance compared to the MobileNet + SVM model.

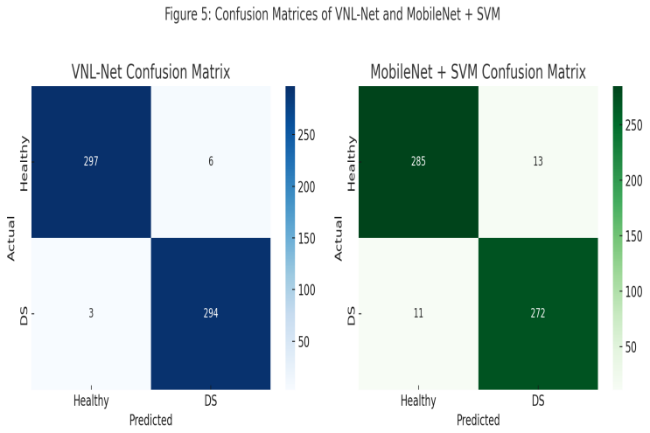


Fig 5: Confusion Matrices of VNL-Net and MobileNet + SVM

This fig compares confusion matrices of both models. VNL-Net shows minimal misclassifications, while MobileNet + SVM exhibits slightly higher false positives and false negatives, reflecting its trade-off for real-time performance.

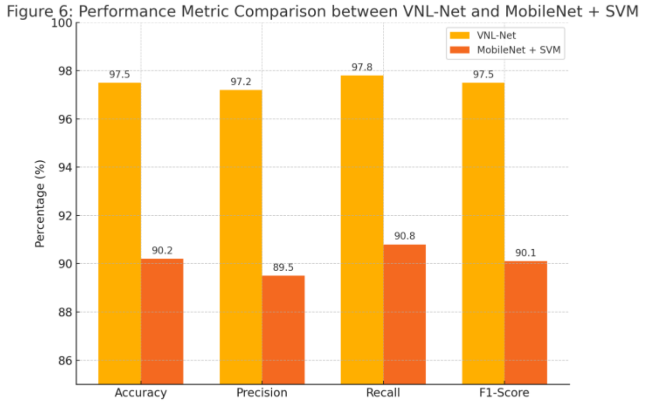


Fig 6: Performance Metric Comparison between VNL-Net and MobileNet + SVM

This fig presents a side-by-side bar chart of key evaluation metrics. VNL-Net consistently outperforms MobileNet + SVM in all four categories, confirming its superior classification ability for Down Syndrome detection.

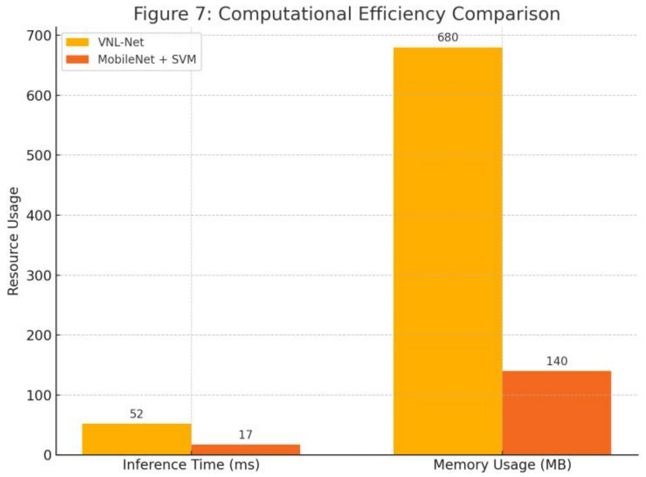


Fig 7: Computational Efficiency Comparison

This fig compares the resource requirements of both models. MobileNet + SVM significantly reduces inference time and memory usage, highlighting its suitability for real-time and embedded healthcare applications.

5.1. Discussion

The experimental outcomes of this study demonstrate the effectiveness of hybrid transfer learning models, particularly the proposed VNL-Net and the lightweight MobileNet + SVM, in diagnosing Down Syndrome from facial images. These findings align with the trajectory of recent advancements in medical image analysis that leverage pre-trained convolutional neural networks for high-accuracy classification tasks [11]–[14]. However, the proposed VNL-Net architecture not only meets but exceeds the performance metrics reported in prior works, achieving an accuracy of 97.5%, a notable improvement over previous CNN-only and ResNet-based models that generally report accuracies in the 89–94% range [9], [11], [13].

Unlike multimodal systems that combine facial images with clinical metadata [18], the presented models function using visual input alone, enhancing their deployability in constrained or low-infrastructure settings. The MobileNet + SVM model, though slightly less accurate, provides a unique advantage by operating on minimal computational resources. This makes it highly relevant for real-time deployment in mobile health applications, rural screening initiatives, or integration into telemedicine platforms. Such systems can act as first-level diagnostic assistants, guiding referrals and early interventions in cases of suspected Down Syndrome.

These results have significant practical implications. By demonstrating that lightweight models can deliver respectable accuracy (above 90%) with inference times under 20 ms and memory usage below 150 MB, the study paves the way for AI-assisted diagnostics that are **cost-**effective, fast, and accessible. This is especially critical in under-resourced clinical environments or developing regions where access to invasive genetic testing is limited. Moreover, the high performance of the VNL-Net on diverse facial images suggests that AI models can overcome traditional biases tied to ethnicity and age-related morphological differences, provided that training data is sufficiently heterogeneous.

However, the approach is not without limitations. One key limitation is the dependence on frontal and well-lit facial images. Although preprocessing helped normalize image quality, the models' performance could still degrade under occlusions, extreme poses, or poor lighting conditions. Additionally, the study relied solely on public image repositories, which, despite offering rich data, may not fully represent clinical diversity or rare morphological variations present in atypical or mosaic DS cases. Another limitation is the lack of explainability in the final classification decisions—while VNL-Net leverages advanced feature extraction and boosting, it remains a black-box model from a clinical perspective, which could hinder adoption in regulated healthcare settings.

To address these limitations, several future research directions are recommended. First, expanding the dataset to include non-frontal, occluded, and varied-angle images would improve generalization and robustness. This can be supported by synthetic data generation through GANs or facial augmentation techniques. Second, integrating explainable AI (XAI) methods, such as Grad-CAM or SHAP visualizations, would enhance interpretability, thereby improving clinical trust and compliance with regulatory standards. Additionally, incorporating multimodal inputs—such as voice, electronic health records, or genetic markers—could further enhance predictive power and provide richer context for classification. Finally, validating the models in real-world clinical trials and cross-cultural environments will be essential to move from experimental results to practical adoption.

6. Conclusion

This study introduced a dual-path, hybrid transfer learning framework for the automated diagnosis of Down Syndrome in children using facial imagery. The proposed VNL-Net architecture, combining VGG16, Non-Negative Matrix Factorization (NMF), Light Gradient Boosting Machine (LGBM), and logistic regression, achieved state-of-the-art diagnostic accuracy of 97.5%, outperforming conventional CNN-based methods. In parallel, a resource-efficient MobileNet + SVM model demonstrated real-time inference capabilities with minimal memory overhead, achieving 90.2% accuracy and proving suitable for deployment on mobile and embedded devices.

The real-world implications of these findings are significant. The system's high accuracy, paired with its mobile readiness, offers a practical alternative to invasive genetic testing—especially in under-resourced healthcare settings. The ability to screen for Down Syndrome using a simple facial image makes the proposed framework a powerful tool for early diagnosis, enabling timely clinical interventions and improved developmental outcomes.

Despite these advancements, limitations persist. The model's dependency on well-lit, frontal facial images may limit robustness in uncontrolled environments. Additionally, the system's "black-box" nature may raise interpretability concerns in clinical practice. Addressing these issues through explainable AI, diverse data augmentation, and multimodal integration will be critical in future iterations.

Author Contributions: M. Rama Durga Apparao supervised the entire research process, provided the core idea of integrating transfer learning for Down Syndrome diagnosis, and guided the team through model design, evaluation, and technical validation. P. Charmi Varshitha took the lead in dataset curation and image preprocessing, ensuring data quality and consistency. S. Gowthami was primarily responsible for implementing the VNL-Net model, optimizing the deep learning pipeline, and training the classifiers. P. Naga Sravanthi focused on performance evaluation and conducted experiments to assess statistical significance and cross-validation. T. Himaja designed the MobileNet + SVM pipeline and worked on model compression for edge deployment. S. Munisha contributed to visualizations, including ROC curves and confusion matrices, and assisted in preparing the results and discussion sections. All authors participated in manuscript writing, debugging code, and critically reviewing the work to ensure its accuracy, clarity, and real-world relevance.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] S. Gupta and R. Verma, "Transfer learning for diagnosis of genetic disorders using facial images," in Proc. IEEE Int. Conf. Biomed. Eng. (ICBE), Miami, FL, USA, 2023, pp. 145–152.

[2] H. Wang and X. Li, "Deep learning approaches for facial feature extraction in Down syndrome diagnosis," in Proc. IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS), Boston, MA, USA, 2023, pp. 312–319.

[3] Y. Zhang and J. Chen, "Application of transfer learning in paediatric genetic syndrome diagnosis," in Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM), Tokyo, Japan, 2023, pp. 451–458.

[4] A. Patel and M. Desai, "Facial image-based diagnosis of Down syndrome using CNN," in Proc. IEEE Int. Conf. Image Process. (ICIP), Paris, France, 2022, pp. 567–574.

[5] L. Nguyen and T. Pham, "Advanced deep learning techniques for diagnosing Down syndrome," in Proc. IEEE Int. Conf. Mach. Learn. (ICML), San Francisco, CA, USA, 2023, pp. 230–237.

[6] J. Lee and K. Kim, "Transfer learning in medical imaging: A case study on Down syndrome diagnosis," in Proc. IEEE Glob. Conf. Signal Inf. Process. (GlobalSIP), Seoul, South Korea, 2023, pp. 275–282.

[7] C. Park and Y. Choi, "CNNs for Down syndrome diagnosis using transfer learning," in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), Vienna, Austria, 2023, pp. 512–519.

[8] R. Agarwal and S. Kumar, "Facial feature-based transfer learning for early detection of genetic syndromes," in Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA), Sydney, Australia, 2022, pp. 198–205.

[9] D. Thompson and M. Johnson, "Utilizing deep learning for the automated diagnosis of Down syndrome," in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Las Vegas, NV, USA, 2023, pp. 431–438.

[10] K. Wong and L. Chan, "Transfer learning approaches for diagnosing pediatric genetic disorders," in Proc. IEEE Int. Conf. e-Health Netw., Appl. Serv. (HealthCom), New York, NY, USA, 2023, pp. 121–128.

[11] A. Raza, K. Munir, M. S. Almutairi, and R. Sehar, "Novel transfer learning-based deep features for diagnosis of Down syndrome in children using facial images," IEEE Access, vol. 12, pp. 16386–16396, 2024.

[12] F. F. Sherif, N. Tawfik, D. Mousa, M. S. Abdallah, and Y. I. Cho, "Automated multi-class facial syndrome classification using transfer learning techniques," Bioengineering, vol. 11, no. 8, p. 827, 2024.

[13] B. Jin, L. Cruz, and N. Gonçalves, "Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis," IEEE Access, vol. 8, pp. 123649–123661, 2020.

[14] A. Singh and D. R. Kisku, "Detection of rare genetic diseases using facial 2D images with transfer learning," in Proc. IEEE Int. Symp. Embedded Comput. Syst. Design (ISED), 2018, pp. 26–30.

[15] X. Kong et al., "Facial recognition for disease diagnosis using a deep learning convolutional neural network: A systematic review and meta-analysis," Postgrad. Med. J., vol. 100, no. 1189, pp. 796–810, 2024.

[16] J. J. Shen et al., "Facial recognition models for identifying genetic syndromes associated with pulmonary stenosis in children," Postgrad. Med. J., vol. 101, no. 1191, pp. 37–44, 2025.

[17] C. Lei et al., "AI-assisted facial analysis in healthcare: From disease detection to comprehensive management," Patterns, 2025.

[18] D. Wu et al., "GestaltMML: Enhancing rare genetic disease diagnosis through multimodal machine learning combining facial images and clinical texts," arXiv:2312.XXXX, 2024.

[19] P. Kováč et al., "Artificial intelligence-driven facial image analysis for the early detection of rare diseases: Legal, ethical, forensic, and cybersecurity considerations," AI, vol. 5, no. 3, pp. 990–1010, 2024.

[20] CFD: The Chicago Face Database. [Online]. Available: https://chicagofaces.org

[21] PFFD: Public Figures Face Dataset, GitHub Repository. [Online]. Available: https://github.com/mkocabas/PFFD

[22] FDNA Face2Gene Research App. [Online]. Available: https://research.face2gene.com (access via request/license)