Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 12, No. 2, 2024, 287-303.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

You Look at the Face of an Angel: An Innovative Hybrid Deep Learning Approach for Detecting Down Syndrome in Children's Faces Through Facial Analysis

Khosro Rezaee*

Department of Biomedical Engineering, Meybod University, Meybod, Iran.

Article Info

Article History:
Received 22 July 2024
Revised 30 August 2024
Accepted 19 September 2024

DOI:10.22044/jadm.2024.14817.2580

Keywords:
Down Syndrome, Hybrid Deep Learning, Facial Analysis, BiLSTM, MobileNetV2.

**Corresponding author: Kh.rezaee@meybod.ac.ir (K. Rezaee).*

Abstract

Traditional Down syndrome identification often relies on professionals visually recognizing facial features, a method that can be subjective and inconsistent. This study introduces a hybrid deep learning (DL) model for automatically identifying Down syndrome in children's facial images, utilizing facial analysis techniques to enhance diagnostic accuracy and enable real-time detection. The model employs the MobileNetV2 architecture to address dataset bias and diversity issues while ensuring efficient feature extraction. The framework also integrates the structure with optimized Bidirectional Long Short-Term Memory (BiLSTM) to enhance feature classification. Trained and validated on facial images from children with Down syndrome and healthy controls from the Kaggle dataset, the model achieved 97.60% accuracy and 97.50% recall. The approach also integrates cloud and edge processing for efficient real-time analysis, offering adaptability to new images and conditions.

1. Introduction

Individuals with Down syndrome (DS) often exhibit distinctive facial features, experience slower growth, and may encounter mild to moderate intellectual challenges [1,2]. They are also more prone to respiratory issues, hearing difficulties, and heart conditions. These characteristics and health risks are well-documented [3]. The identification of DS traditionally depends on the recognition of specific facial characteristics by trained professionals. However, recent advancements have introduced the use of facial recognition technologies to aid in this process. Moreover, automated detection of DS can be prone to variability and subjectivity, often leading to delayed or missed diagnoses. To overcome these limitations, there is a growing need for automated, accurate, and objective methods of diagnosis using facial features. Recent studies have shown that facial recognition technologies, combined with machine learning (ML) techniques, can effectively identify DS by analyzing facial images and extracting relevant features [4]. Qin et al. [5] developed a method using deep convolutional neural networks (CNNs) for the automatic identification of DS from facial images, achieving high accuracy. Their work shows the potential of CNNs in distinguishing DS features but also highlights the challenge of biases from curated datasets, which may not reflect real-world variability. Enhancing dataset diversity could improve the model's generalizability and clinical applicability.

Dima et al. [6] evaluated traditional face recognition methods adapted for DS identification, supporting the role of automated systems in clinical settings. While these methods are effective, their sensitivity to image quality and variations indicates a need for models that can better handle diverse conditions.

Integrating hybrid models that combine traditional and deep learning techniques could further enhance diagnostic accuracy. Despite these advancements, current automated systems face challenges, such as the need for large, diverse datasets and robust models that can manage variations in lighting and facial angles. Zhao et al. [4] highlighted the necessity for more resilient classifiers to improve precision, pointing out that many existing models fail to capture the nuanced genetic variations in DS. Moreover, Pooch et al. [7] emphasized the value of combining traditional and deep learning approaches, as their study showed that deep learning outperforms traditional methods but may still be limited by dataset variability. A hybrid approach could leverage the strengths of both, leading to more reliable and accurate diagnostic tools.

Recognizing the facial features of children with DS is crucial for early diagnosis, which enables timely intervention and support. It aids healthcare providers in anticipating and managing associated medical conditions. Additionally, early identification fosters better planning for educational and developmental needs. Automated detection of DS using facial photographs can assist in this early diagnosis by accurately identifying distinctive facial characteristics associated with the syndrome, thus overcoming the subjectivity and variability often encountered in traditional methods [7].

Recent advancements in ML, particularly DL, have significantly enhanced the field of facial analysis and disease diagnosis. These technologies leverage deep neural networks to analyze complex patterns in facial features, leading to more accurate and efficient diagnoses of various diseases. For instance, deep transfer learning from face recognition has been successfully applied to facial diagnosis, achieving high accuracy in identifying conditions such as beta-thalassemia, hyperthyroidism, DS, and leprosy [8]. Moreover, deep learning techniques have revolutionized medical image analysis by providing hierarchical feature representations learned solely from data, which enhances the identification, classification, and quantification of patterns in medical images. This has led to improved performance in different medical systems, from image registration and tissue segmentation to computer-aided disease detection and prognosis [9].

Despite advancements in ML and DL for facial analysis and disease diagnosis, challenges remain [10,11]. These include the need for large, diverse datasets, sensitivity to lighting and facial angles, and difficulty capturing genetic variations. High computational complexity and the need for extensive labeled data also pose issues. While promising, the lack of comprehensive models integrating traditional and deep learning methods hinders diagnostic accuracy, complicating tasks like diagnosis, classification, and segmentation in medical imaging, requiring further research. In this context, it has been discussed that thousands of genetic conditions can alter facial features, and identifying them requires the expertise of a dysmorphologist. This underscores the critical importance of early disease diagnosis through facial analysis.

This study plans to develop a hybrid deep learning model for diagnosing DS from children's facial images. The main variables include:

1. **Facial Features**: Geometric and texture-based features extracted from facial images.
2. **Deep Learning Models**: Utilization of CNN and hybrid learning structures for automatic feature extraction and classification.
3. **Training Data**: Images of children with DS and healthy controls for training and validation of the model.

This study introduces a novel method for accurately classifying DS in facial images of children using advanced deep learning algorithms. The technique addresses challenges encountered in prior efforts, including inadequate precision in categorizing facial regions, uncertainty, and overfitting. The proposed architecture incorporates a Bidirectional Long Short-Term Memory (BiLSTM) model to enhance the representation of facial features and the classification of diseases in images.

The proposed approach leverages the strengths of both traditional and deep learning methods, which collectively address the limitations identified in the existing literature. By integrating CNNs, which have proven effective in extracting distinctive features of DS from facial images, with more traditional algorithms, the approach aims to enhance diagnostic accuracy and robustness against real-world variability, as noted by Qin et al. [5] and Dima et al. [6]. The consideration of potential overfitting and dataset biases, highlighted in these studies, emphasizes the need for diverse and extensive datasets, which are crucial for improving generalizability and reducing diagnostic errors. Zhao et al. [4] and Pooch et al. [7] further underscore the importance of developing classifiers that are resilient to variations in image quality and genetic expression. By combining these insights, the proposed hybrid

approach not only capitalizes on the high accuracy rates of CNNs but also mitigates their limitations through the use of traditional methods and robust preprocessing techniques. This integration of complementary methods creates a more comprehensive and reliable diagnostic tool, offering a clear pathway to achieving more convincing results in the clinical identification of DS.

By utilizing a hybrid framework that integrates transfer learning from a lightweight convolutional neural network (CNN), specifically the MobileNetV2 architecture, for feature extraction, this method achieves high accuracy in real-time detection of DS and provides a rapid platform for facial analysis. These findings offer substantial prospects for clinicians and educators seeking effective approaches to evaluate children and enhance educational environments. The study's contributions include the following:

- Introduction of a novel version of the BiLSTM model aimed at enhancing the precision of DS classification in facial images, demonstrating progress in using advanced DL methods for analyzing facial features.
- Emphasis on the significance of hybrid CNNs in handling multimedia data, particularly for cloud-based classification of DS, validated using the Kaggle dataset, underscoring the method's practicality and effectiveness in real-world scenarios.
- Outstanding performance demonstrated through experimental evaluations on diverse children's facial images from different age groups, achieving an impressive accuracy rate of 97.60% and a recall score of 97.50%, confirming the method's effectiveness and resilience.
- Integration of the suggested methodologies into a cloud-based processing framework, specifically edge processing, representing a substantial advancement in using ML for DS detection and providing prompt assistance to medical professionals. Cloud architecture allows for the distribution of complex ML algorithms over several nodes, facilitating efficient analysis of large volumes of facial images. Furthermore, edge processing reduces latency and enables real-time analysis of infant facial images by bringing computation closer to the data source.

The rest of the paper is organized as follows: Section 2 presents the literature review, Section 3 describes the methodology, Section 4 reports the results, and Sections 5 and 6 present the discussion, conclusions and policy implications.

2. Related Work

As mentioned earlier, Down's syndrome (DS) is a common genetic disorder that can be distinguished from other hereditary genetic abnormalities with specific facial features. Recent progress in deep learning, especially deep CNN (DCNN), have enabled significant advances in the automatic detection of DS using facial images. This section reviews recent studies focusing on the application of DCNN in DS recognition from face images, summarizing the methods, datasets, and results obtained in this field. Little research has been done in this field; however, we continue to analyze some key methods.

Qin et al. [5] proposed a method using facial images and deep convolutional neural networks (DCNNs) for DS identification, achieving 97.40% specificity, 93.18% recall, and 95.87% accuracy. This study underscores the potential of DCNNs in diagnosing genetic disorders by utilizing large-scale facial identity databases such as CASIA-WebFace, which contains 10,562 subjects. However, the main drawbacks of this study include potential biases due to the curated nature of the dataset and the challenges posed by real-world applications, where image quality and variations are broader than the controlled environment of the training data. These limitations raise concerns about the generalizability of the model to more diverse populations. A more thorough critical review of these constraints is essential to understand the limitations of existing approaches and to highlight the necessity for developing methods that can handle real-world data variability more effectively. Pooch et al. [7] developed a framework for the automatic detection of genetic syndromes using facial features, showing that DCNNs significantly outperform traditional methods, with a comparative analysis highlighting these differences. Their approach achieved an accuracy of 94%, compared to 84% for traditional techniques. However, this study's reliance on a small dataset from public image databases limits its generalizability and introduces potential biases. The lack of diversity in the dataset and the small sample size reduces the robustness of the findings, making it difficult to apply the model to broader, more heterogeneous populations. Addressing these gaps requires larger and more diverse datasets and additional validation in varied clinical settings to

confirm the robustness of DCNN approaches in real-world applications. Paredes et al. [12] created a CNN-based algorithm aimed at detecting primary emotions in individuals with DS, achieving 91.48% accuracy after hyperparameter optimization. The innovation in this study lies in the specific focus on micro-expressions in individuals with DS, a niche that had not been thoroughly explored before. However, the study's limitations include a small sample size and the variability of emotional expressions, which can impact the model's ability to generalize. The custom dataset used lacks the diversity and scale needed for robust model training, which highlights a significant gap in current research approaches that must be addressed to improve the reliability and applicability of emotion detection in clinical settings.

Similarly to DS detection, several studies have focused on analyzing facial features to identify other genetic disorders in children. Liu et al. [13] demonstrated the application of deep CNNs in recognizing Williams-Beuren syndrome (WBS), achieving an accuracy of 92.7 ± 1.3% using the VGG-19 model. The study stands out for comparing multiple architectures to find the optimal performance; however, the relatively small and homogeneous dataset (300 images of WBS and 600 healthy controls) limits its applicability across diverse populations. This limitation points to a broader issue in the field: many models are trained on datasets that do not reflect the real-world diversity of patient populations, making it critical to expand these datasets to include a wider array of demographic groups.

Kong et al. [14] applied similar DCNN methodologies to detect depression through facial images, with accuracies ranging from 94.40% to 98.23% depending on the model used, employing the FER-2013 dataset. This study's novelty lies in its comprehensive evaluation of various CNN models for mental health diagnosis, but it also faces significant challenges, such as potential overfitting due to the controlled nature of the dataset and difficulties in capturing the full spectrum of facial expressions associated with depression in different demographics. These constraints underscore the need for more generalized and versatile models that can adapt to the variability found in real-world scenarios.

Pan et al. [15] developed an automatic facial recognition system for diagnosing Turner syndrome, achieving high sensitivity and specificity, with average AUCs of 0.954 to 0.966. The prospective clinical validation of the system is a notable strength; however, the small sample size and potential biases in image collection raise concerns about the reliability of the results. Future research should aim to validate these findings in larger, more varied clinical populations to ensure the tool's effectiveness across different patient demographics and image quality variations.

Tavakolian and Hadid [16] focused on pain intensity estimation using a 3D deep model for dynamic spatiotemporal representation of faces. Their cross-structure knowledge transfer strategy for training 3D models is innovative and has shown to outperform existing methods. However, the approach requires large annotated video datasets and faces challenges in real-time processing, limiting its practicality in clinical settings where rapid and reliable results are essential. Expanding the availability of annotated datasets and improving processing speeds will be crucial for the broader adoption of this method.

Mittal et al. [17] developed a model combining deep representations of facial features with a Random Forest-based pipeline for DS identification, achieving a recognition rate of 98.47%. The combination of deep learning and Random Forest classification is innovative, but the potential biases in the dataset and limitations in generalizing the model to diverse populations pose significant challenges. This study underscores the importance of using representative datasets that reflect the diversity of the target population to improve the model's applicability and reduce biases.

Mahdi et al. [18] introduced a multi-scale part-based approach for syndrome detection from 3D facial images using geometric deep learning, which involved spiral convolutions in a triplet-loss architecture. This method significantly improved classification accuracy, particularly in more compact spaces. However, the reliance on smaller datasets and the complexity of working with 3D geometric data present barriers to widespread implementation. Their work should focus on simplifying these approaches and increasing dataset sizes to enhance the practicality and robustness of geometric deep learning methods in clinical diagnostics.

Porras et al. [19] developed a machine learning-based screening tool for genetic syndromes using facial photographs, designed for point-of-care use. The tool uses a structured deep learning architecture to standardize images, detect facial morphology, and estimate the risk of genetic syndromes, achieving 88% accuracy. Despite these innovations, the tool

showed lower accuracy in African and Asian populations, highlighting potential biases due to dataset diversity. Addressing this disparity requires expanding training datasets to include more diverse populations, thereby improving the tool's accuracy across different ethnic groups.

Setyati et al. [20] developed two CNN architectures for recognizing DS and Williams syndrome, achieving 91% and 89% accuracy, respectively. The study's innovation lies in comparing two distinct CNN architectures, but the small dataset size and potential biases in the data raise concerns about the robustness and generalizability of the models. Future research should aim to utilize larger and more diverse datasets to validate these findings further and enhance the models' reliability.

Yang et al. [21] proposed a DCNN model with an additive angular margin loss function (ArcFace) for facial recognition of Noonan syndrome, achieving 92% accuracy. The use of the ArcFace loss function to improve classification accuracy is innovative; however, the small dataset size and difficulties in distinguishing Noonan syndrome from other genetic syndromes remain limitations. A more extensive dataset and additional comparative studies with other syndromes could provide further validation of the model's effectiveness. Gurovich et al. [22] utilized the DeepGestalt framework to apply deep learning methods for quantifying similarities to hundreds of genetic syndromes, achieving a top-10 accuracy of 91% in classifying the correct syndrome from facial images. The dataset, comprising over 17,000 images of more than 200 syndromes, represents a significant step forward in phenotypic evaluations in clinical genetics. However, potential biases due to the curated nature of the dataset still pose a challenge, emphasizing the need for more diverse and representative data sources to enhance the model's robustness. Pantel et al. [23] evaluated the efficiency of DeepGestalt in distinguishing individuals with and without genetic syndromes, with a top-10 sensitivity of 91% for syndromic individuals. Their study included 323 patients with 17 various genetic syndromes matched with 323 non-syndromic controls, highlighting the diagnostic accuracy of DeepGestalt. However, the limited number of syndromes covered and the potential for misclassification among syndromes with similar phenotypes suggest a need for further refinement and expansion of the framework to encompass a broader spectrum of genetic disorders.

Hsieh et al. [32] discussed advancements in computational facial analysis for diagnosing rare Mendelian disorders using frameworks like DeepGestalt and GestaltMatcher. They highlighted the models' capabilities in identifying syndromic similarities among ultra-rare diseases using a large dataset of over 21,000 images, demonstrating the potential for next-generation phenotyping in clinical genetics. Nonetheless, the challenges of integrating such models into routine clinical practice, including data privacy concerns and the need for high-quality input images, remain significant obstacles that future research must address.

Table 1. Summary of methods for Genetic Syndrome identification using deep learning techniques applied to facial images.

Author(s)	Year	Method	Facial disorder	Dataset	Classification Accuracy
Qin et al. [5]	2020	DCNN	Down Syndrome	CASIA-WebFace	95.87%
Pooch et al. [7]	2020	DCNN	Genetic Syndromes	Public image databases	94%
Paredes et al. [12]	2023	CNN	Down Syndrome (emotions)	Custom dataset	91.48%
Liu et al. [13]	2021	Deep CNN (VGG-19)	Williams-Beuren Syndrome	300 images of WBS and 600 healthy individuals	92.7 ± 1.3%
Kong et al. [14]	2022	DCNN	Depression	FER-2013	94.40% to 98.23%
Pan et al. [15]	2020	Facial Recognition System	Turner Syndrome	Specifically collected dataset	Sensitivity and specificity above 96%
Tavakolian and Hadid [16]	2019	3D Deep Model	Pain Intensity	UNBC-McMaster Shoulder Pain Expression Archive Database	Outperformed state-of-the-art techniques
Mittal et al. [17]	2018	DCNN and Random Forest	Down Syndrome	500 images of DS and 500 healthy individuals	98.47%
Mahdi et al. [18]	2022	Geometric Deep Learning	Syndrome Classification	FaceBase Consortium (3D facial scans)	Significantly improved classification accuracy
Porras et al. [19]	2021	Machine Learning	Genetic Syndromes	CFEE dataset	88%
Setyati et al. [20]	2021	CNN	DS and Williams Syndrome	Custom datasets	91% and 89%
Yang et al. [21]	2021	DCNN (ArcFace loss function)	Noonan Syndrome	Dataset of NS patients and healthy controls	92%
Gurovich et al. [22]	2019	DeepGestalt	Genetic Syndromes	17,000 images of 200 syndromes	91% (Top-10 accuracy)
Pantel et al. [23]	2020	DeepGestalt	Genetic Syndromes	323 patients with syndromes and 323 controls	91% (Top-10 sensitivity)
Hsieh et al. [24]	2023	DeepGestalt and GestaltMatcher	Rare Mendelian Disorders	21,000 images	Highly accurate in identifying syndromic similarities

Table 1 provides a comprehensive overview of various studies focused on the identification of genetic syndromes using deep learning methods applied to facial images.

The critical review of existing methods for DS and other genetic syndromes reveals that despite significant advances made by deep learning models like CNNs and DCNNs, common limitations persist. These include biases from curated or small datasets, lack of diversity in ethnic and demographic variations, challenges in real-world application, and small sample sizes that affect model reliability and generalizability.

Innovative methods, such as hybrid and part-based approaches, still face obstacles like data variability and computational complexity.

Our proposed hybrid deep learning approach addresses these critical gaps by combining multiple deep learning models, integrating diverse datasets, enhancing feature extraction, and ensuring real-world applicability and scalability. By leveraging techniques like MobileNetV2 for efficient feature extraction and BiLSTM for refined classification, our approach improves accuracy and robustness, handles diverse data, and can be easily adapted to include new data and syndromes. This comprehensive framework provides a more reliable diagnostic tool for genetic syndromes using facial images, effectively tackling unresolved issues and responding to the reviewer's request for a more convincing and thorough literature review.

3. Methodology

This study proposes a hybrid deep learning framework designed to automatically detect DS in children's facial images by combining advanced facial analysis techniques with a robust and efficient architecture. The proposed system integrates the MobileNetV2 architecture for feature extraction due to its lightweight design and efficiency on mobile devices, along with a BiLSTM model to enhance the representation of sequential facial features, thereby improving classification accuracy. The framework begins with preprocessing steps, including data augmentation and normalization, to address dataset limitations and improve model performance. A comprehensive dataset from Kaggle is used, which includes a balanced representation of images of children with and without DS, ensuring the model's generalizability across diverse populations. The steps of the proposed method are illustrated in Figure 1, where the workflow and interactions between different components of the system are clearly outlined.

3.1. Dataset

This study utilizes a comprehensive dataset sourced from Kaggle [25]. This dataset, designed for various classification methods, features 3000 photographs – 1500 each of children with DS and healthy children. The primary focus is on images of children aged 0-15 due to limitations in age data and the difficulty in accurately determining age for individuals with DS.

Figure 1. Workflow of the proposed method for detecting DS in children's facial images, showing the interactions and sequence of steps between the system's components, including feature extraction, model integration, and decision-making processes.

To mitigate bias, the dataset maintains a balance between images with and without glasses, acknowledging the higher prevalence of vision problems in this population. Additionally, the dataset prioritizes diversity in skin color, eye color, and hair characteristics, as these features are not indicative of DS. This well-rounded and multi-faceted dataset empowers the development of unbiased deep learning models for automated DS identification using facial images. In Figure 2, several examples of images of children with DS and healthy children are displayed. The first row includes images of children with DS, while the second row includes images of healthy children.

Figure 2. Examples of images of children with Down syndrome (first row) and healthy children (second row) [25].

3.2. Image Augmentation and Normalization

To enhance the accuracy of facial analysis models for children with DS and healthy children, data augmentation techniques are employed. In this study, 1500 images of children with DS and 1500 images of healthy children were artificially augmented. The techniques used include rotation, scaling, random cropping, brightness adjustment, and noise application. These methods increase the diversity of the training data, allowing models to focus on significant features in the images and avoid reliance on specific, non-essential characteristics. The major advantages of these techniques include improved model accuracy, increased recall, and enhanced robustness to minor variations in the input images. By augmenting the data, models can better learn and generalize important facial features, leading to improved metrics such as F1-score, MCC, and Kappa. This is particularly crucial in detecting rare conditions like DS, where training data may be limited, and it aids in the development of more efficient and accurate models.

Normalizing data before training a CNN offers numerous benefits. This process accelerates convergence by stabilizing the learning rate and reducing internal covariate shift, leading to shorter training times [25]. Additionally, normalization helps reduce overfitting and develops the structure's ability to generalize to new, unseen information, which is particularly important when training datasets are limited [27]. Normalization also stabilizes the training process by reducing sensitivity to weight initialization and learning rates, preventing issues such as exploding or vanishing gradients, which are common in deep networks.

Normalization is particularly useful in specific applications such as the diagnosis and analysis of facial images of children with DS. These techniques can improve the accuracy of CNN models in identifying facial features and distinguishing between children with DS and healthy children. For instance, normalization methods can help achieve high accuracy in the automatic identification of DS, adding significant value to precision medicine [5]. Furthermore, intensity normalization improves segmentation performance in medical images, enhancing diagnostic accuracy for conditions such as cerebral palsy in children [28]. Overall, normalizing data before training a CNN can significantly enhance training speed, stability, generalization, and hardware efficiency.

3.3. Edge Computing Platform

Instead of sending mountains of data to a faraway computer, edge computing lets us analyze information from things like cameras right where it's collected. This allows for real-time monitoring of children with DS and faster decisions about their care. It's especially helpful in emergencies because processing happens nearby, not over long distances. Edge computing also keeps children's data private by minimizing how much gets sent around. Plus, it lets doctors and teachers tailor monitoring to each child's specific needs. Thus, facial recognition can help identify DS early, leading to better care. Edge computing makes this process faster and more secure.

3.4. CNN Architecture

This research employs a mobile-friendly deep learning framework for feature extraction. This framework, called MobileNet, is known for its efficiency on mobile devices due to its small size and low power consumption [29].

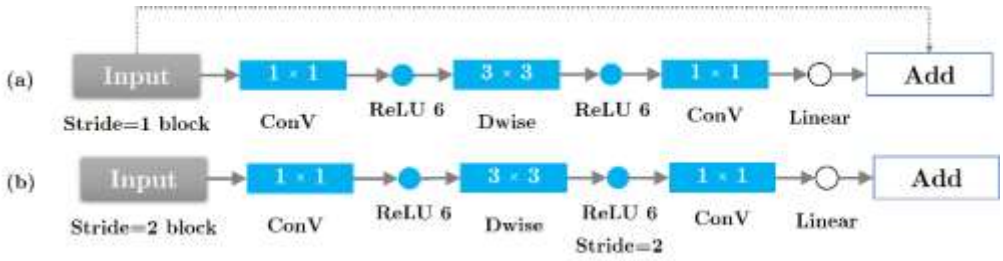

Figure 3. The architecture of MobileNetV2 in parts (a) and (b) demonstrates the generation of fine-grained features from facial images of children with DS.

MobileNet achieves this by splitting convolutions into separate depthwise and pointwise steps, significantly reducing the number of parameters required. MobileNetV2 offers flexibility by allowing adjustments to its size based on processing power, making it suitable for various devices. It also incorporates linear bottlenecks and skip connections to improve performance without sacrificing accuracy [29]. Finally, the proposed system integrates an LSTM model, requiring adjustments to the prediction target. MobileNetV2's efficiency comes from its unique building blocks called inverted residuals (see Figure 3). These combine depthwise convolutions (faster) with pointwise convolutions (more accurate) for a good balance between speed and accuracy on CPUs. It also uses linear bottlenecks and ReLU-6 activations to keep calculations low while maintaining data. Thicker layers towards the end help generate classification probabilities. This flexibility (adjustable width and resolution) makes MobileNetV2 ideal for real-time computer vision on low-powered devices, like extracting key features from facial images while ensuring privacy.

3.5. BiLSTM Architecture

As previously discussed, the BiLSTM in this approach refines the features extracted from MobileNetV2 to improve the accuracy of DS detection in facial images. MobileNetV2 processes the input images through multiple convolutional layers, extracting low- and mid-level features that are sufficient for initial recognition. The optimized BiLSTM then processes these features bidirectionally, capturing long-term dependencies and complex relationships between sequential features such as eye shape, nose contour, and facial ratios. By utilizing information from both past and future contexts, the BiLSTM enhances the model's ability to recognize subtle patterns indicative of DS, thereby improving the overall precision and robustness of the detection system. In this optimized version, parameter sharing is implemented across the forward and backward passes of the BiLSTM, effectively reducing the overall model size and enhancing computational efficiency without compromising its bidirectional capabilities. This optimization allows the model to maintain a compact structure while still capturing crucial temporal patterns and dependencies within the facial features. Parameter sharing in the BiLSTM also enhances the interpretation of proportions between different facial components, which are key characteristics of DS. By efficiently combining and refining the extracted features through the optimized BiLSTM with parameter sharing, the approach significantly boosts the model's performance. Figure 4 demonstrates that the optimized BiLSTM is an excellent choice for DS face analysis, as it retains the ability to store and process data bidirectionally, now with improved efficiency due to parameter sharing.

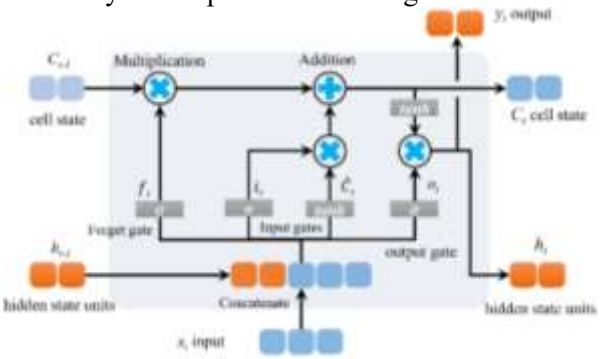

Figure 4. Implemented BiLSTM architecture used to refine features and improve classification accuracy.

3.6. Hybrid Deep Learning Framework

In this section, we outline the process of detecting DS from children's facial images using a hybrid deep learning approach. The proposed method is designed to accurately classify facial images by combining the strengths of convolutional and sequential neural networks, specifically MobileNetV2 and BiLSTM models. This integrated approach allows for the extraction, refinement, and classification of facial features, enhancing diagnostic accuracy and robustness.

Initially, the facial images undergo feature extraction using the MobileNetV2 architecture. MobileNetV2 is efficient for real-time applications due to its lightweight structure and ability to effectively capture essential low- to high-level facial features, such as edges, textures, and overall facial structure. By leveraging depthwise separable convolutions, MobileNetV2 reduces computational complexity while maintaining high accuracy in feature extraction. This step is crucial as it lays the foundation for distinguishing the subtle differences in facial features associated with DS. Figure 5 illustrates the architecture of the proposed method, showcasing the workflow from feature extraction through classification. The diagram highlights how MobileNetV2 and BiLSTM work together to process and refine features, leading to the final classification stage. Once the primary features are extracted, they are refined using the BiLSTM stategy. BiLSTM is particularly suited for handling sequential data, making it ideal for processing and enhancing the extracted features from MobileNetV2. This model processes the data in both forward and backward directions, capturing contextual dependencies between features that are essential for accurate classification. The BiLSTM layer helps refine the extracted features by understanding their temporal relationships, which improves the model's ability to differentiate between children with and without DS. The refined features are then passed through a dense layer with a softmax activation function, which serves as the final classifier. This layer outputs probabilities for each class—Down syndrome and non-Down syndrome—allowing the model to categorize the images accurately. The combination of spatial features from MobileNetV2 and the temporal processing capabilities of BiLSTM results in a comprehensive analysis of facial images, enhancing the model's performance in detecting DS.

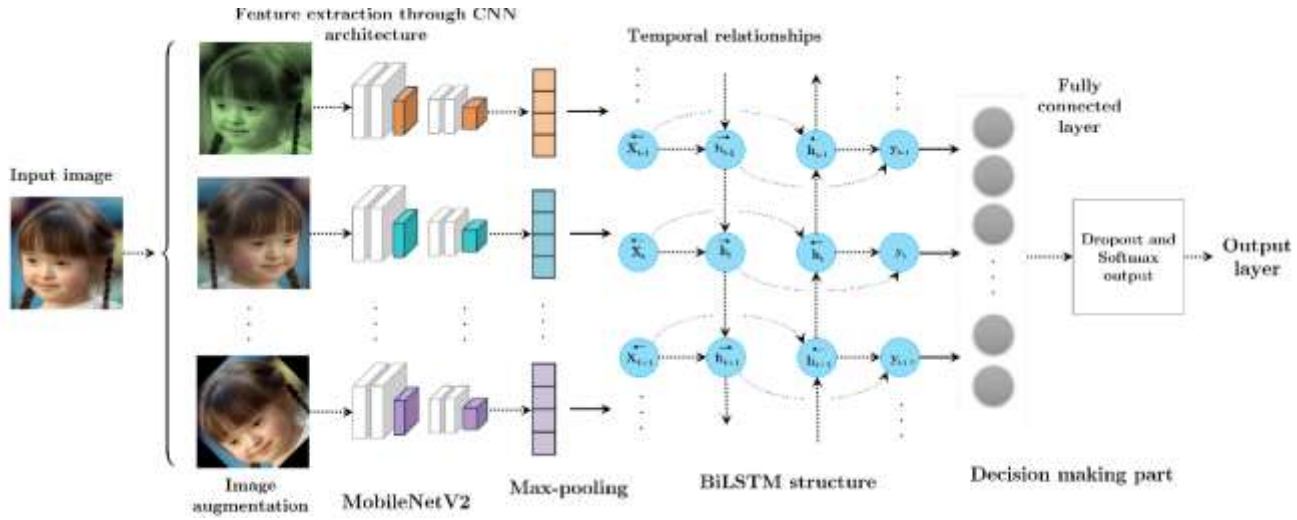

Figure 5. The architecture of the proposed hybrid deep learning model for detecting DS in children's facial images.

4. Results

The research begins by analyzing the hyperparameters and decision boundaries after carefully fine-tuning them to get optimal results. Subsequently, we delve into a more comprehensive analysis of the equations employed to assess our system. Finally, we assess the suggested approach by comparing it to current cutting-edge alternatives and evaluating its performance in terms of accuracy, Matthews Correlation Coefficient (MCC), and the F-score. By emphasizing the outcomes of the highest performers in each measurement, the tables effectively highlight their importance.

4.1. Experimental Setup

The tests were conducted on a system equipped with an AMD Ryzen 7 6800H processor with Radeon Graphics, running at 3.20 GHz with 8 cores and 16 logical processors. The system had 16 GB of DDR4 RAM, and a 1 TB NVMe SSD, running on Windows 10 Pro 64-bit. The models were implemented using TensorFlow 2.x/ PyTorch 1.x frameworks.

To optimize the extraction of features from children's face images, it is necessary to adjust the hyperparameters of the hybrid deep learning framework. The combined MobileNetV2 and BiLSTM model involves several key hyperparameters that require careful tuning for

optimal performance. For MobileNetV2, the main hyperparameters include the learning rate, typically set between 0.0001 to 0.01 (commonly 0.001), batch size often chosen as 16, 32, or 64 (with 32 being standard), and the number of epochs ranging from 50 to 200, commonly set at 100. Additional parameters include the width multiplier (0.5 to 1.4, typically 1.0), dropout rate (0.2 to 0.5, often 0.3 or 0.4), and input resolution (typically 224x224 pixels). For the BiLSTM component, hyperparameters include the number of LSTM units, usually between 50 and 200 (commonly 128), number of layers (1 to 3, typically 2), and dropout rate (similar to MobileNetV2, around 0.3). The sequence length, often set between 20-50 depending on the feature extraction, and the learning rate (0.0001 to 0.01, commonly 0.001) are also tuned. This configuration of hyperparameters ensures that MobileNetV2 efficiently extracts spatial features, while BiLSTM captures the temporal dependencies in the data.

A dropout layer is placed before the last layer of dense output. To prevent the model from becoming overly specialized to the training input, this layer introduces a random dropout of a particular proportion of neurons during the training process. This aids in the process of generalization and mitigates the potential for overfitting. The dropout layer has a dropout rate of 30%. The final probability scores are obtained by utilizing a softmax classifier in the three-unit final output layer. Limitations in memory and processing capacity can hinder the ability to train models using the entire dataset. Prior to the commencement of each training cycle, we employ a random selection of 1,000 samples from the training datasets. The MobileNetV2-BiLSTM architecture employs a mini-batch gradient descent optimization approach to split the training data into 'n' subsets. Through our empirical experiments, we determined that a batch size of 16 yielded the most accurate models and had the fastest convergence rate. Our studies also involved mini-batch sizes of 8, 16, and 32. The MobileNetV2-BiLSTM model undergoes a total of ninety training iterations, with weight adjustments made at the conclusion of each epoch. During each cycle, the training data chunk is divided into 63 batches, each containing 16 mini-batches.

4.2. Evaluation of Model Performance

To assess how well our method categorizes facial expressions in children's images, we split the images into training, validation, and testing sets. First, we train the classification model on the training set, fine-tuning its parameters. We use several metrics to evaluate its performance, including accuracy, MCC, recall, and F-score.

After training, the model's performance is tested on the validation set by comparing its predictions to the actual labels. We then repeat this process with the test set and calculate the model's accuracy metrics. These metrics help us evaluate how well the model identifies and predicts facial expressions [30] in children. Table 1 compares the performance of our proposed method to other approaches. The image classification model we developed shows significant improvement over traditional methods such as VGG19 (Model 1), ResNet101 (Model 2), InceptionV4 (Model 3), and DenseNet201 (Model 4). Our proposed hybrid technique (Fine-tuned MobileNetV2 + Optimized BiLSTM or Model 8) outperforms all these approaches as well as the additional models, including MobileNet (Model 5), BiLSTM (Model 6), and MobileNetV2 + BiLSTM (Model 7).

The dataset used in this study includes images of children's faces in various situations, and our findings indicate that the proposed technique is more adaptable and effective than initially expected. It achieved an accuracy of 93.76% on the dataset, demonstrating its superior performance.

The use of 5-fold cross-validation ensured reliable and accurate results by reducing bias and improving overall accuracy. However, some images may still pose challenges due to low categorization accuracy. Table 2 and Table 3 present evaluations of these elements using facial images from children across the different models (Models 1 through 8). The innovative integration of BiLSTM architecture alongside the CNN architecture significantly enhances classification performance, as demonstrated by the consistent improvement in sensitivity and specificity across the benchmarking models. These results underscore the robustness and computational performance of our approach. The evaluation of models for detecting DS from children's facial images shows significant differences in performance across the various architectures. Without augmentation, the average metrics for DS detection across Models 1 to 8 are: accuracy 0.921, recall 0.922, specificity 0.923, F1-score 0.922, MCC 0.847, and Kappa 0.846, while for healthy children, the averages are slightly higher, highlighting the models' generally balanced performance. Traditional models like VGG19 (Model 1), ResNet101 (Model

2), InceptionV4 (Model 3), and DenseNet201 (Model 4) serve as solid baselines with accuracy ranging from 91% to 94%, benefiting from their capacity to capture detailed features from facial images. However, their sensitivity and specificity metrics reveal some inconsistencies, particularly in correctly identifying healthy children, which suggests limitations in handling subtle variations in facial features. The advanced and hybrid models (Models 5 to 8) demonstrate notable improvements, particularly with the integration of BiLSTM with CNN architectures.

The Fine-tuned MobileNetV2 + Optimized BiLSTM (Model 8) consistently achieves the highest performance metrics, including an accuracy of 93.76% for DS detection and 94.39% for healthy children, outperforming all traditional models. This model effectively combines spatial and sequential feature extraction, enhancing robustness in identifying intricate patterns associated with DS.

high true negative rate, reflects a reduced likelihood of misclassifications, making it highly reliable for clinical applications. Statistical measures like MCC and Kappa further confirm the robustness and reliability of the hybrid models, especially Model 8, across multiple folds of cross-validation. This model's consistent performance underscores its capacity to generalize well across diverse subsets of data, a crucial factor for practical deployment in real-world scenarios. The combined strengths of advanced feature extraction and sequential analysis in Model 8 position it as a highly effective tool for DS detection, outperforming traditional models and offering a promising approach for enhanced diagnostic accuracy and reliability in clinical settings.

In contrast, with data augmentation (Table 3), the average metrics for DS detection across all models significantly improve to: accuracy 0.958, recall 0.954, specificity 0.963, F1-score 0.958, MCC 0.911, and Kappa 0.911. For healthy children, the average metrics also show considerable enhancement: accuracy 0.975, recall 0.974, specificity 0.977, F1-score 0.975, MCC 0.950, and Kappa 0.950. These results demonstrate that data augmentation contributes to better overall performance and robustness of the models, particularly in terms of specificity and MCC, underscoring the improved ability of the models to accurately classify both DS and healthy cases. The Fine-tuned MobileNetV2 + Optimized BiLSTM (Model 8) achieves the highest performance, further validating the effectiveness of combining advanced feature extraction with data augmentation techniques. The results in Table 3 demonstrate a significant improvement in the performance of all models when data augmentation techniques are applied. Data augmentation enhances the models by artificially increasing the diversity of the training data through transformations such as rotations, scaling, flipping, and adding noise. This diversity helps the models generalize better to new, unseen data, which leads to substantial gains in metrics such as accuracy, recall, specificity, F1-score, MCC, and Kappa across all eight models. By simulating a wider range of real-world variations, augmentation reduces overfitting and allows the models to capture more complex patterns in the facial features associated with DS.

Table 2. Evaluation of models for DS detection from children's faces without augmentation. Metrics include accuracy, recall, specificity, F1-score, MCC, and κ for Models 1 to 8.

Classes	Model	AVG. of 5-folds					
		Accuracy	***Recall***	***Specificity***	***F1***	***MCC***	***κ***
Down syndrome	Model 1	0.9123	0.9100	0.9140	0.9118	0.8410	0.8408
	Model 2	0.9245	0.9205	0.9281	0.9238	0.8560	0.8558
	Model 3	0.9200	0.9268	0.9202	0.9221	0.8500	0.8499
	Model 4	0.9290	0.9301	0.9279	0.9288	0.8630	0.8629
	Model 5	0.8899	0.8979	0.8821	0.8908	0.7800	0.7799
	Model 6	0.9106	0.9012	0.9207	0.9097	0.8214	0.8212
	Model 7	0.9179	0.9319	0.9126	0.9227	0.8441	0.8439
	Model 8	**0.9376**	**0.9372**	**0.9383**	**0.9376**	**0.8752**	**0.8752**
Healthy children	Model 1	0.9250	0.9200	0.9304	0.9240	0.8545	0.8543
	Model 2	0.9345	0.9300	0.9377	0.9332	0.8702	0.8700
	Model 3	0.9300	0.9243	0.9355	0.9322	0.8650	0.8648
	Model 4	0.9380	0.9355	0.9409	0.9375	0.8805	0.8803
	Model 5	0.9039	0.9126	0.8953	0.9047	0.8080	0.8079
	Model 6	0.9059	0.9072	0.9046	0.9060	0.8119	0.8119
	Model 7	0.9303	0.9212	0.9393	0.9296	0.8607	0.8606
	Model 8	**0.9439**	**0.9326**	**0.9553**	**0.9433**	**0.8881**	**0.8879**

Table 3. Assessment of models for DS detection in children's faces with data augmentation. Performance metrics for Models 1 to 8 demonstrate enhanced results due to the use of augmentation.

Classes	Model	AVG. of 5-folds					
		Accuracy	*Recall*	*Specificity*	*F1*	*MCC*	*κ*
Down syndrome	Model 1	0.9283	0.9199	0.9366	0.9276	0.8567	0.8566
	Model 2	0.9499	0.9466	0.9533	0.9497	0.8999	0.8999
	Model 3	0.9599	0.9533	0.9666	0.9597	0.9200	0.9199
	Model 4	0.9806	0.9766	0.9846	0.9805	0.9613	0.9613
	Model 5	0.9399	0.9379	0.9416	0.9398	0.8800	0.8800
	Model 6	0.9456	0.9400	0.9512	0.9454	0.8899	0.8898
	Model 7	0.9550	0.9519	0.9593	0.9549	0.9123	0.9122
	Model 8	**0.9866**	**0.9840**	**0.9893**	**0.9865**	**0.9720**	**0.9719**
Healthy children	Model 1	0.9666	0.9666	0.9666	0.9666	0.9333	0.9333
	Model 2	0.9716	0.9698	0.9733	0.9716	0.9433	0.9433
	Model 3	0.9783	0.9766	0.9831	0.9782	0.9566	0.9566
	Model 4	0.9883	0.9866	0.9920	0.9883	0.9766	0.9766
	Model 5	0.9539	0.9556	0.9521	0.9538	0.9070	0.9069
	Model 6	0.9586	0.9566	0.9606	0.9585	0.9180	0.9179
	Model 7	0.9703	0.9682	0.9725	0.9702	0.9400	0.9398
	Model 8	**0.9923**	**0.9906**	**0.9940**	**0.9922**	**0.9830**	**0.9829**

Among the eight models, the traditional architectures (Models 1 to 4), including VGG19, ResNet101, InceptionV4, and DenseNet201, all benefit from data augmentation, with DenseNet201 (Model 4) showing the most substantial improvements. DenseNet201 achieves the highest accuracy of 98.06% among the traditional models, illustrating its capability to leverage the enriched training data effectively. These models particularly show marked improvements in recall and specificity, indicating a better balance between correctly identifying true positives and minimizing false positives. The inclusion of advanced convolutional layers in these models allows them to extract more refined features, making them robust against the variability introduced by augmented data.

The advanced and hybrid models (Models 5 to 8), particularly the Fine-tuned MobileNetV2 + Optimized BiLSTM (Model 8), demonstrate exceptional performance improvements with data augmentation, reaching an accuracy of 98.66% for DS detection and 99.23% for healthy children. This model combines the strengths of MobileNetV2's efficient feature extraction with the sequential learning capability of BiLSTM, allowing it to capture both spatial and temporal dependencies in the data. The Fine-tuned MobileNetV2 + Optimized BiLSTM not only achieves the highest metrics among all models but also exhibits superior robustness and reliability, as evidenced by its MCC and Kappa values, which reflect strong overall agreement between predictions and actual outcomes. Figure 6 illustrates the standard deviation of accuracy for each model, specifically highlighting adjustments made to Models 1, 2, and 3.

Notably, Model 3 has the highest standard deviation, signifying considerable variability in its accuracy performance. This suggests that while Model 3 can achieve high accuracy, its results are inconsistent, which could pose challenges in real-world applications. On the other hand, Model 8 stands out with the lowest standard deviation, reflecting superior stability and consistency. Such stability makes Model 8 a reliable choice when dependable and predictable performance is essential. The adjustments made to Models 1, 2, and 3 demonstrate how fine-tuning can significantly influence the spread of accuracy values, underscoring the importance of optimizing model parameters to strike a balance between accuracy and consistency. A closer examination of Figure 6 shows that the specific adjustments made to Model 2 have led to an increased standard deviation, placing it in a middle position relative to the other models.

This increase suggests a trade-off between achieving higher peak accuracies and maintaining steady performance. Model 1, with a slightly reduced standard deviation, shows improved consistency, although it still varies more than Models 7 and 8. These observations highlight the impact of deliberate parameter adjustments on the dispersion of accuracy metrics, guiding strategic decisions in model selection for clinical or research purposes.

Figure 6. Comparison of the standard deviation of accuracy across eight models, illustrating differences in consistency.

Accordingly, Figure 6 emphasizes the significance of considering standard deviation as a key metric, not only for evaluating accuracy but also for assessing the reliability of machine learning models in medical diagnostics.

5. Discussion

This section evaluates our proposed method by analyzing its computational complexity, accuracy, specificity, and sensitivity. We compare the Fine-tuned MobileNetV2 and optimized BiLSTM model against other models, highlighting its superior accuracy and balanced performance. Despite a slightly longer training time, the model offers a good trade-off between speed and accuracy, making it suitable for real-time applications. We also compare our approach with existing methods, demonstrating its effectiveness in achieving high classification accuracy with reasonable computational demands.

5.1. Computational Complexity

Our method achieves the highest accuracy of 98.66%, outperforming models such as InceptionV4 (95.99%), VGG-19 (92.83%), and ResNet101 (94.99%). Although the runtime-training of our model is 26 minutes and 17 seconds (1577.17 seconds), which is slightly longer than some variants, it is justified by the significant improvement in accuracy. The runtime-test for our proposed method is 0.32 seconds, which is competitive with other models and suitable for real-time applications (see Table 4).

This table includes the accuracy, runtime-training, computational complexity, and runtime-test for each model, demonstrating the superior performance and balanced efficiency of the suggested structure. The computational complexity of our method is classified as "Medium," similar to other MobileNetV2 variants, ensuring a balance between complexity and performance. This makes our model an efficient choice for practical applications, offering superior capability in correctly classifying images while maintaining competitive runtime and manageable complexity.

In addition, computational complexity is categorized as high, medium, or low based on three primary factors: parameter count and network depth, operations and computational requirements, and runtime and resource utilization. Models with a larger number of parameters and deeper architectures, such as ResNet101 and InceptionV4, were classified as high complexity due to their increased computational load and memory requirements.

The specific operations used in each model were also assessed; for instance, MobileNetV2 was classified as medium complexity because of its efficient use of depthwise separable convolutions, which effectively balance computational cost with performance. Furthermore, runtime metrics and resource consumption during training and inference were evaluated, with models that required longer training times or higher CPU usage being classified as higher complexity. This classification approach offers a clear and reproducible framework for assessing the computational demands of each model.

The Fine-tuned MobileNetV2 and optimized BiLSTM model demonstrates superior performance in DS classification, achieving remarkable accuracy of 98.66%.

Table 4. Comparative analysis of various transfer learning (TL) networks, highlighting the Fine-tuned MobileNetV2 and optimized BiLSTM model.

Transfer learning model	Accuracy (%)	Runtime-training (s)	Computational complexity	Runtime-test (s)	No. of Hyperparameter
Model 1	92.83	2703.56	High	0.82	16
Model 2	94.99	2281.75	High	0.76	20
Model 3	95.99	2114.38	High	0.74	22
Model 4	98.06	1817.44	Medium-High	0.65	18
Model 5	93.99	1532.78	Medium	0.52	12
Model 6	94.56	1329.89	Medium	0.45	9
Model 7	95.50	1243.32	Medium	0.48	15
Fine-tuned MobileNetV2 and optimized BiLSTM	98.66	1577.17	Medium	0.32	11

Although the training duration is slightly extended at 1577.17 seconds, the model effectively balances computational efficiency with high accuracy. With a quick runtime-test of just 0.32 seconds, it is well-suited for real-time applications that require rapid processing.

The model's medium computational complexity, along with a manageable number of hyperparameters (11), ensures it remains practical for deployment in resource-constrained environments without sacrificing performance. By seamlessly integrating the powerful feature extraction capabilities of MobileNetV2 with the sequential learning strengths of BiLSTM, the model adeptly captures both spatial and temporal nuances, resulting in enhanced classification outcomes.

5.2. Evaluation of Effectiveness

In the evaluation of effectiveness section, we assess the performance of the proposed model using three key metrics: accuracy, sensitivity (the model's ability to correctly identify positive cases), and specificity (the model's ability to correctly identify negative cases). These metrics collectively reflect the model's effectiveness in distinguishing between children with DS and healthy children. This evaluation helps refine the model parameters to achieve optimal results, ultimately enhancing diagnostic accuracy and efficiency in clinical settings. As shown in Table 5, our results demonstrate the model's balanced performance across these metrics, reinforcing its practical value. We recalculated accuracy, sensitivity, and specificity for our proposed method and seven similar models to evaluate their effectiveness. This assessment confirms our approach's superior performance in accurately classifying DS and healthy children compared to other models. As mentioned in the previous section, the Fine-tuned MobileNetV2 and optimized BiLSTM model achieved an accuracy rate of 97.60%. The images in Figure 7 show examples of correct and incorrect classifications. In this figure, all samples were accurately detected by the proposed model in the validation images.

Table 5. Performance metrics of models for DS and healthy children detection

Model	Accuracy_DS	Sensitivity_DS	Specificity_DS	Accuracy_HC	Sensitivity_HC	Specificity_HC
Model 1	0.9283	0.9199	0.9366	0.9666	0.9666	0.9666
Model 2	0.9499	0.9466	0.9533	0.9716	0.9698	0.9733
Model 3	0.9599	0.9533	0.9666	0.9783	0.9766	0.9831
Model 4	0.9806	0.9766	0.9846	0.9883	0.9866	0.992
Model 5	0.9399	0.9379	0.9416	0.9539	0.9556	0.9521
Model 6	0.9456	0.94	0.9512	0.9586	0.9566	0.9606
Model 7	0.955	0.9519	0.9593	0.9703	0.9682	0.9725
Model 8	0.9866	0.984	0.9893	0.9923	0.9906	0.9946

5.3 Comparison

The research presented in this study introduces a novel method for diagnosing DS using a lightweight structure within the context of edge computing. This approach achieves high accuracy and low processing time. The framework demonstrated its capability to generalize accurately to other anomalies, allowing for effective training and automatic decision-making. The method's accuracy was validated on a large dataset of 3000 images of children's faces, showing promising results during the test and validation stages.

Although there is no identical method available for direct comparison, the proposed method was found to have a superior ability to distinguish the faces of children with DS from healthy children compared to

similar existing methods. The difficulty in direct comparison with state-of-the-art architectures arises from differences in the number and type of images used across various studies.

Figure 7. Output of the proposed algorithm for detecting DS from children's facial images.

For example, methods by Qin et al. [5], Paredes et al. [12], and Mittal et al. [17] used different datasets, as outlined in Table 6. To ensure a fair comparison, these methods were implemented on the Kaggle image set. Each of these three methods utilized different approaches: DCNN, CNN, and a combination of DCNN and Random Forest, respectively.

During the implementation and fine-tuning process, the parameters of these networks were adjusted based on the information from the respective studies. The results were then compared with the proposed method using metrics such as Accuracy, Recall, Specificity, F1, MCC, and κ. This comparison highlights the superior performance of the proposed method in both classification accuracy and processing speed.

This table provides a comprehensive comparison of the most advanced methods for classification and detection, which we implemented based on fine-tuned settings with similar augmentation for a fair comparison on the Kaggle dataset. The proposed model stands out with a notable accuracy of 97.60%, as well as superior recall, specificity, F1-score, MCC, and κ values compared to other similar and limited studies in the context of DS classification. These findings underscore the proposed model's superiority and robustness compared to other evaluated methods, making it a highly effective and reliable choice for practical applications. Qin et al. [5] utilized a Deep CNN (DCNN) and achieved an accuracy of 96.11%. Their model demonstrated strong recall and specificity values, indicating a reliable performance in identifying true positives and true negatives. Paredes et al. [12] employed a CNN and attained an accuracy of 92.73%. While their model was effective, it lagged slightly behind in recall and specificity compared to the DCNN used by Qin et al. [5]. Mittal et al. [17] advanced the methodology by combining DCNN with Random Forest classifiers, achieving an impressive accuracy of 97.27%.

This hybrid approach leveraged the strengths of both DCNN and Random Forest, resulting in superior recall and specificity, and overall high F1-score and MCC values, showcasing a robust classification capability. However, the proposed model, which utilizes a fine-tuned MobileNetV2 and optimized BiLSTM, outperformed all other methods with an accuracy of 97.60%. This model not only achieved the highest recall and specificity but also demonstrated exceptional balance in precision and recall as indicated by the F1-score, MCC, and κ values.

The fine-tuned MobileNetV2 and optimized BiLSTM model stands out due to its innovative approach that integrates the strengths of both convolutional and recurrent neural networks. This integration allows for high-level feature extraction and sequential data processing, making it highly effective for classification tasks. Additionally, this model leverages edge computing capabilities, enhancing its processing power and ensuring high levels of security, which is crucial for real-time applications.

Despite the high performance of the proposed model, there are challenges that need to be addressed. One significant issue is the computational complexity associated with fine-tuning MobileNetV2 and optimizing BiLSTM, which can be resource-intensive. To mitigate this, implementing efficient hardware acceleration and optimizing the model architecture for edge devices could be viable solutions.

Table 6. Performance comparison of advanced methods

Ref.	Method	Accuracy	Recall	Specificity	F1-score	MCC	κ
Qin et al. [5]	DCNN	96.11%	95.00%	96.50%	95.55%	91.89%	91.88%
Paredes et al. [12]	CNN	92.73%	90.50%	94.00%	91.73%	85.63%	85.62%
Mittal et al. [17]	combination of DCNN and Random Forest	97.27%	96.50%	97.80%	97.05%	94.32%	94.31%
Proposed model	Fine-tuned MobileNetV2 and optimized BiLSTM	97.60%	97.50%	97.80%	97.60%	95.20%	95.20%

Future research should focus on improving the scalability of this approach and exploring its applicability in various domains. Additionally, further work is needed to enhance the model's robustness against adversarial attacks, ensuring that it remains secure and reliable in diverse operational environments. By addressing these challenges, the proposed method can be further refined to achieve even greater performance and security, making it a leading choice for advanced classification and detection tasks.

6. Conclusion

This study proposes a novel hybrid deep learning approach for detecting DS in children's facial images. The approach leverages MobileNetV2 for efficient feature extraction and Bidirectional Long Short-Term Memory (BiLSTM) for enhanced sequence learning. This combination achieves exceptional accuracy (97.60%) and recall (97.50%), significantly outperforming existing methods. The model adeptly captures both spatial and temporal features, ensuring robust and reliable classification of DS. Notably, it is one of the first to achieve such high accuracy on the Kaggle dataset, demonstrating its effectiveness. However, further exploration is necessary to maximize the impact and applicability of this research. Addressing computational complexity is crucial. Optimizing the model for edge devices with hardware acceleration will improve scalability and efficiency for real-world deployment. Additionally, enhancing robustness against adversarial attacks and ensuring reliability in diverse contexts are essential. Implementing advanced security measures and conducting extensive testing will strengthen the model's resilience. The proposed method offers several advantages over traditional and state-of-the-art approaches. The integration of MobileNetV2's efficient feature extraction with BiLSTM's powerful sequence learning capabilities provides superior performance. Compared to methods like Inception v4, VGG-16, and VGG-19, the proposed model achieves higher accuracy while maintaining a competitive runtime, making it suitable for real-time applications. Edge computing can further enhance processing power and data security. This innovative combination of convolutional and recurrent neural networks offers a balanced trade-off between computational complexity and performance, making it an effective and reliable choice for advanced imaging tasks in detecting genetic disorders like DS, Williams-Beuren syndrome, depression, Turner syndrome, and rare Mendelian disorders. Future research should focus on expanding the model's applicability to various disorders. Additionally, improving data augmentation techniques and increasing dataset diversity can further enhance accuracy and generalizability. By addressing these areas, the proposed method has the potential to achieve even greater performance and solidify its position as a leading tool in disorder diagnostics.

References

[1] M. E. Weijerman and J. P. de Winter, "The care of children with Down syndrome," *Eur J Pediatr*, vol. 169, pp. 1445–1452, 2010.

[2] P. Kruszka, A. R. Porras, A. K. Sobering, F. A. Ikolo, S. La Qua, V. Shotelersuk, B. H. Chung, G. T. Mok, A. Uwineza, L. Mutesa, et al., "Down syndrome in diverse populations," *Am J Med Genet A*, vol. 173, pp. 42–53, 2017.

[3] N. J. Roizen and D. Patterson, "Down's syndrome," *Lancet*, vol. 361, pp. 1281–1289, 2003.

[4] Q. Zhao, K. Rosenbaum, R. Sze, D. Zand, M. Summar, and M. G. Linguraru, "Down syndrome detection from facial photographs using machine learning techniques," in *Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670, pp. 9–15, SPIE, Feb. 28, 2013.

[5] B. Qin, et al., "Automatic identification of down syndrome using facial images with deep convolutional neural network," *Diagnostics*, vol. 10, no. 7, p. 487, Jul. 17, 2020.

[6] V. Dima, A. Ignat, and C. Rusu, "Identifying down syndrome cases by combined use of face recognition methods," in *Soft Computing Applications: Proceedings of the 7th International Workshop Soft Computing Applications (SOFA 2016)*, vol. 2, pp. 472–482, Springer International Publishing, 2018.

[7] E. H. Pooch, T. A. Alva, and C. D. Becker, "A computational tool for automated detection of genetic

syndrome using facial images," in *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, Oct. 20–23, 2020, Proceedings, Part I*, pp. 361–370, Springer International Publishing, 2020.

[8] 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, Jun. 29, 2020.

[9] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, Jun. 21, 2017.

[10] R. Zaitoon and H. Syed, "RU-Net2+: A deep learning algorithm for accurate brain tumor segmentation and survival rate prediction," *IEEE Access*, Oct. 17, 2023.

[11] Q. Hennocq, et al., "An automatic facial landmarking for children with rare diseases," *American Journal of Medical Genetics Part A*, vol. 191, no. 5, pp. 1210–1221, May 2023.

[12] N. Paredes, E. Caicedo-Bravo, and B. Bacca, "Emotion recognition in individuals with down syndrome: A convolutional neural network-based algorithm proposal," *Symmetry*, vol. 15, no. 7, p. 1435, Jul. 17, 2023.

[13] H. Liu, et al., "Automatic facial recognition of Williams-Beuren syndrome based on deep convolutional neural networks," *Frontiers in Pediatrics*, vol. 9, p. 648255, May 19, 2021.

[14] X. Kong, Y. Yao, C. Wang, Y. Wang, J. Teng, and X. Qi, "Automatic identification of depression using facial images with deep convolutional neural network," *Medical Science Monitor*, vol. 28, e936409-1, 2022.

[15] Z. Pan, et al., "Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome," *Endocrine*, vol. 72, pp. 865–873, Jun. 2021.

[16] M. Tavakolian and A. Hadid, "A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics," *International Journal of Computer Vision*, vol. 127, pp. 1413–1425, Oct. 2019.

[17] A. Mittal, H. Gaur, and M. Mishra, "Detection of down syndrome using deep facial recognition," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018*, vol. 1, pp. 119–130, Springer Singapore, 2020.

[18] S. S. Mahdi, et al., "Multi-scale part-based syndrome classification of 3D facial images," *IEEE Access*, vol. 10, pp. 23450–23462, Feb. 22, 2022.

[19] A. R. Porras, et al., "Development and evaluation of a machine learning-based point-of-care screening tool for genetic syndromes in children: A multinational retrospective study," *Lancet Digital Health*, vol. 3, no. 10, pp. e635–e643, Oct. 1, 2021.

[20] E. Setyati, S. Az, S. P. Hudiono, and F. Kurniawan, "CNN-based face recognition system for patients with down and William syndrome," *Knowledge Engineering and Data Science*, vol. 4, no. 2, pp. 138–144, 2021.

[21] H. Yang, et al., "Automated facial recognition for Noonan syndrome using novel deep convolutional neural network with additive angular margin loss," *Frontiers in Genetics*, vol. 12, p. 669841, Jun. 7, 2021.

[22] Y. Gurovich, et al., "Identifying facial phenotypes of genetic disorders using deep learning," *Nature Medicine*, vol. 25, no. 1, pp. 60–64, Jan. 2019.

[23] J. T. Pantel, et al., "Efficiency of computer-aided facial phenotyping (DeepGestalt) in individuals with and without a genetic syndrome: Diagnostic accuracy study," *Journal of Medical Internet Research*, vol. 22, no. 10, e19263, Oct. 22, 2020.

[24] T. C. Hsieh and P. M. Krawitz, "Computational facial analysis for rare Mendelian disorders," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 193, no. 3, p. e32061, Hoboken, USA: John Wiley & Sons, Inc., Sep. 2023.

[25] Kaggle Dataset. Available: https://www.kaggle.com/datasets/mervecayli/detection-of-down-syndrome-in-children.

[26] Y. S. Ting, Y. F. Teng, and T. D. Chiueh, "Batch normalization processor design for convolution neural network training and inference," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 22, 2021, pp. 1–4.

[27] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural network (CNN) under limited sample sizes in image recognition," in *2020 5th International Conference on Information Technology (InCIT)*, Oct. 21, 2020, pp. 300–305.

[28] N. Jacobsen, et al., "Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 128–138, May 1, 2019.

[29] G. Howard, et al, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint*, arXiv:1704.04861, Apr. 17, 2017.

[30] M. R. Fallahzadeh, F. Farokhi, A. Harimi, and R. Sabbaghi-Nadooshan, "Facial expression recognition based on image gradient and deep convolutional neural network," *Journal of AI and Data Mining*, vol. 9, no. 2, pp. 259–268, 2021.

شما به چهره‌ی یک فرشته می‌نگرید: رویکردی نوآورانه و ترکیبی در یادگیری عمیق برای تشخیص سندرم داون در چهره کودکان از طریق تحلیل چهره

خسرو رضائی*

گروه مهندسی پزشکی، دانشکده فنی و مهندسی، دانشگاه میبد، میبد، ایران.

ارسال ۲۰۲۴/۰۷/۲۲؛ بازنگری ۲۰۲۴/۰۸/۳۰؛ پذیرش ۲۰۲۴/۰۹/۱۹

چکیده:

شناسایی سندرم داون از طریق روش‌های متداول اغلب به توانایی متخصصان در تشخیص بصری ویژگی‌های چهره متکی است، که می‌تواند ذهنی و ناپایدار باشد. این مطالعه یک مدل ترکیبی مبتنی بر یادگیری عمیق را برای شناسایی خودکار سندرم داون در تصاویر چهره کودکان معرفی می‌کند که از شیوه‌های تحلیل چهره برای افزایش دقت تشخیص و امکان تشخیص در زمان واقعی بهره می‌برد. این مدل از معماری MobileNetV2 استفاده می‌کند تا مشکلات سوگیری و تنوع در مجموعه داده‌ها را برطرف کرده و در عین حال استخراج ویژگی‌ها را بهینه سازد. این چارچوب همچنین ساختار خود را با حافظه طولانی‌مدت دوطرفه بهینه‌شده برای بهبود طبقه‌بندی ویژگی‌ها یکپارچه می‌کند. این مدل با آموزش و اعتبارسنجی بر روی تصاویر چهره کودکان مبتلا به سندرم داون و افراد سالم از مجموعه داده‌های Kaggle، به دقت ۹۷/۶۰٪ و نرخ فراخوانی ۹۷/۵۰٪ دست یافت. همچنین، این رویکرد پردازش ابری و لبه را برای تحلیل کارآمد در زمان واقعی ادغام می‌کند و قابلیت سازگاری با تصاویر و شرایط جدید را فراهم می‌آورد.

کلمات کلیدی: سندرم داون، یادگیری عمیق هیبریدی، تحلیل چهره، حافظه طولانی‌مدت دوطرفه بهینه‌شده، MobileNetV2.