Journal of Intelligent Systems and Internet of Things
Vol. 15, No. 01, PP. 133-143, 2025
DOI: https://doi.org/10.54216/JISIoT.150111

Deep Learning for Multi-Label Facial Attribute Classification on Large-Scale Image Datasets (CelebA)

Sif. K. Ebis[1], Bushra Majeed Muter[1], Fatima Hameed Shnan[1], Oday Ali Hassen[1,*]
[1]Ministry of Education, Wasit Education Directorate, Iraq
Emails: Saifkather@gmail.com; bushramajeed1975@gmail.com; fatemahameed1984@gmail.com; oday123456789.oa@gmail.com

Abstract

The exponential growth of data in recent years has led to an increasing demand for advanced techniques, especially those that work on large and complex data. This has given deep learning a significant advance in dealing with the tasks of analyzing, improving, and distinguishing big data. Our research focused on CNNs from this data and applying deep learning algorithms and their analysis to a large-scale image dataset. More specifically, our research focused on a dataset called CelebA, which contains more than 200,000 face images annotated with 40 binary facial features. It is a multi-label classification model based on the ResNet-50 architecture that has been fine-tuned to predict different facial features and hair color such as age, gender, and facial expressions. It was also trained using data augmentation, taking into account pose differences and background clutter to reduce imbalance between classes. These results reflect very strong predictive performance, with an average mean accuracy of 0.86 and an overall F1 score of 0.81 across all features. Attributes identified by clear visual cues—for example, "smiling," "male "and" wearing lipstick"—were highly accurate, while less obvious attributes such as "big lips" and "narrow eyes" were more difficult to classify. We would like to point out that the results demonstrate the high efficiency of using deep learning models for multi-label classification on big data while solving problems associated with class imbalance and overfitting models. This research leads to the larger general field of big data analytics; in particular, it demonstrates how deep learning can be efficiently applied to large image datasets for automatic attribute recognition. It also opens up potential applications in areas such as biometric identification, surveillance, and human-computer interaction.

Received: July 04, 2024 Revised: September 28, 2024 Accepted: December 26, 2024

Keywords: Deep Learning; Convolutional Neural Networks; Big Data Analysis; Multilabel Classification; Large-scale Clubfaces Attributes (CelebA) Dataset; ResNet-50; Image Classification

1. Introduction

Analytics and extraction of insights from such volumes of information are an essential challenge in business, healthcare, education, or government in this big data-driven world. The exponentially increasing data will be digitally generated through online platforms, sensors, and multimedia in the future, for which advanced processing techniques must be developed to interpret and act on such information. [1][2]. One of the most promising methods is deep learning, an area of machine learning whose major breakthrough has recently shaken up a whole range of fields, including image and speech recognition, natural language processing, and autonomous systems. Deep learning techniques, especially CNNs, are able to extract hierarchical features from raw data in such an exciting manner that they become quite effective in computer vision. [3][4]. This work will therefore be devoted to a review of the performance of deep learning algorithms over large-scale image datasets, considering their application to the CelebA dataset, which includes over 200,000 annotated face images [4][5][6]. The CelebA dataset is recognized universally for its applicability in facial attribute prediction tasks; it is an ideal choice with which to conduct research on the performance of deep learning models in big data scenarios.

The dataset involves 40 binary attributes per image, along with feature types such as gender, hair color, and facial expressions. Multilabel classification problems may explore rich settings within the dataset. This paper has applied a customized deep learning model based on ResNet-50 for multilabel classification, which predicts a set of facial attributes from each image. It is evaluated for its performance with extensive experimentation, including augmentation of data and strategies for addressing class imbalance, hence understanding the strengths and limitations in handling complex high-dimensional data. Results obtained from this research work therefore showcase the applicability of deep learning towards big data analysis on applications that span from biometric identifications and surveillance to advanced human-computer interaction systems [7].

2. Related Work

In [6][8] present a review on the usage of Machine Learning and Deep Learning techniques to perform advanced data analysis. In this paper, the latest developments and applications of ML methods have been reviewed, which includes supervised, unsupervised, and reinforcement learning techniques, while for deep learning it reviews CNNs, RNNs, and transformers. These deep learning structures are truly impressive in their ability to manage complex and high-dimensional data. Some of these key topics will include, but are not limited to: data preparation, feature selection, and noisy or missing data handling. Other topics might involve applications in industries such as health, finance, and retail; scaling challenges in ML/DL models; and current trends, including automated ML, edge computing, and the integration of quantum computing. Ethics around data privacy and bias is discussed.

In [7][9] research investigates Big Data to enhance the performance, accuracy, and scalability of ML. The study describes how volume, variety, and velocity of Big Data have assisted ML models in improving generalization and reducing overfitting, hence allowing the support of complex deep learning architectures. The discussion is done based on the tools that can be used to process Big Data in ML workflows, including Apache Hadoop and Spark. Case studies within healthcare, finance, and social media are provided, followed by benefits and challenges found in the real world. The paper also tries to present computational demands of Big Data and addresses ethical concerns of data privacy. Future research into quantum computing and ensuring fairness within AI systems forms the focus.

Hend A. Selmy and Hoda K. Mohamed's paper [8][21] offers a detailed survey of deep learning (DL) techniques used in big data analytics (BDA). The study emphasizes DL's ability to handle unstructured, large-scale data, making it highly valuable in fields like speech recognition, image classification, and IoT data processing. It provides an overview of key DL techniques, their taxonomy, and their applications in real-world big data scenarios. The survey also compares various DL frameworks and datasets, discussing their strengths and limitations. Finally, it addresses the challenges of DL modeling and outlines future research directions.

Angel Ojeda and Juan Valera [9][22] researched how Big Data and AI are correlated to one another to enhance data analysis in business decision-making. Interestingly, analyzing 3,574 data scientists and analysts across the world using Kaggle made identification of some key patterns between AI techniques and types of data possible. The results indicated that 66% of AI data mining algorithms have been applied to structured data, while 37% of the applications relate to diagnostic analysis of data. Only 3% of the prescriptive data types are mainly analyzed using the text-mining AI algorithms. This finding gives insight into how businesses can better select applicable AI-driven analysis techniques in driving informed decisions.

3. Proposed Methodology

The proposed methodology of this research is to conduct deep learning approaches for multilabel classification on a large-scale facial image dataset, CelebA, consisting of over 200,000 face images which have been annotated for 40 distinct facial attributes.[10][11][23] The primary aim here will be the detection and classification of multiple facial attributes in an image automatically with the use of a CNN model, specifically a ResNet-50 architecture. First, dataset preprocessing was performed: loading and merging of attribute and partition files, conversion of binary labels to an appropriate format, and augmentation techniques to increase the generalization ability of the model.

Also, simple augmentation techniques were applied: random horizontal flipping, random rotation, and resizing to simulate various natural conditions included in the dataset and thus helped avoid some overfitting. We split the dataset accordingly into training, validation, and testing sets according to the provided partitions so that the model is properly validated during training.

The ResNet50 model is adapted for multi-attribute classification through modifications in the final fully connected layers. It facilitates multi-classification with sigmoid activation and thus can utilize BCE loss. Efficient gradient descent was done with the AdamW optimizer, which was combined with a learning rate scheduler for fine-tuning the model on how it learns over time.

The performance would, therefore, be assessed in terms of accuracy, precision, recall, and F1-score. Also, to get an insight into the ability of the model in deeper detail while predicting each attribute, confusion matrices and classification reports will be obtained. Visualization is also included in the proposed methodology, whereby the model's predictions are plotted against actual labels for sample images to show its practical performance on unseen data.

CelebFaces Attributes (CelebA) Dataset CelebA is among the benchmarks that are most widely used in computer vision, considered standard for recognition of facial attributes and other tasks related to it. This dataset, which was collected by researchers from the Chinese University of Hong Kong, has a total of more than 202,599 single images showing celebrity faces that represent a wide range of ethnicities, ages, and facial features. [12][13]

Each image in this dataset is complemented with 40 binary attribute labels that describe the presence of specific characteristics such as gender, hair color, and facial expressions. This extensive labeling enables researchers to train and evaluate deep learning models on multilabel classification tasks, and because of this, CelebA is an invaluable resource for developing algorithms that can understand complex visual information. [14][15]

It also includes bounding box annotations and landmark coordinates for the facial features, further enriching its utility in a range of applications involving face recognition and analysis. Due to the comprehensive nature of this dataset, the CelebA dataset has been widely utilized in research related to facial recognition systems, attribute prediction, and generative modeling and thus stands at the core in developing methodologies in the field.[25]

Table 1: CelebA Attributes

Attribute	Description
5_o_Clock_Shadow	Presence of a five o'clock shadow
Arched_Eyebrows	Eyebrows are arched
Attractive	Subject is considered attractive
Bags_Under_Eyes	Presence of bags under the eyes
Bald	Subject is bald
Bangs	Subject has bangs
Big_Lips	Subject has large lips
Big_Nose	Subject has a prominent nose
Black_Hair	Subject has black hair
Blond_Hair	Subject has blonde hair
Blurry	Image is blurry
Brown_Hair	Subject has brown hair
Bushy_Eyebrows	Eyebrows are bushy
Chubby	Subject is chubby
Double_Chin	Subject has a double chin
Eyeglasses	Subject is wearing eyeglasses
Goatee	Subject has a goatee
Gray_Hair	Subject has gray hair
Heavy_Makeup	Subject is wearing heavy makeup

High_Cheekbones	Subject has high cheekbones
Male	Subject is male
Mouth_Slightly_Open	Subject's mouth is slightly open
Mustache	Subject has a mustache
Narrow_Eyes	Subject has narrow eyes
No_Beard	Subject has no beard
Oval_Face	Subject has an oval face
Pale_Skin	Subject has pale skin
Pointy_Nose	Subject has a pointy nose
Receding_Hairline	Subject has a receding hairline
Rosy_Cheeks	Subject has rosy cheeks
Sideburns	Subject has sideburns
Smiling	Subject is smiling
Straight_Hair	Subject has straight hair
Wavy_Hair	Subject has wavy hair
Wearing_Earrings	Subject is wearing earrings
Wearing_Hat	Subject is wearing a hat
Wearing_Lipstick	Subject is wearing lipstick
Wearing_Necklace	Subject is wearing a necklace
Wearing_Necktie	Subject is wearing a necktie
Young	Subject is considered young

Table 1 summarizes the 40 attributes that provide a comprehensive overview of the characteristics analyzed within the dataset, facilitating the development and evaluation of models for facial attribute recognition.

Proposed ResNet50 Model.

The proposed ResNet50 model is the convolutional neural network to learn representation and extract effective features from complex image data with regard to facial attribute recognition on the CelebA dataset. ResNet50 is an extension of ResNet architecture, which introduced the concept of residual learning through skip connections whereby the network managed to reduce the vanishing gradient problem that is normally faced in deep networks. It has 50 layers, while convolutional layers are merged with batch normalization and ReLU activations that enhance feature extraction and improve convergence during training.

In our implementation, we use a pre-trained ResNet50 model as the backbone; hence, transfer learning takes benefit from knowledge learned on extensive training over large datasets.

We further fine-tune the model for multi-label classification tasks associated with facial attributes by adjusting the final fully connected layer to produce the number of attributes present in the CelebA dataset. Furthermore, the introduction of dropout layers provides further regularization in a manner that reduces overfitting when training the network.

By nature, the architecture of ResNet50 is such that identification and prediction of facial attributes could be done rather robustly by it, as it is very suited for the analysis and recognition for the face.

Table 2: Proposed ResNet50 Model Parameters.

Parameter	Description
Input Size	224 x 224 pixels (for each input image)
Number of Layers	50 layers
Convolutional Layers	49 convolutional layers (including 1x1 and 3x3 filters)
Fully Connected Layers	1 fully connected layer (modified to output the number of attributes)
Batch Normalization	Applied after each convolutional layer to normalize activations
Activation Function	ReLU (Rectified Linear Unit) used as the activation function in hidden layers
Dropout Rate	0.3 (applied before the final fully connected layer to reduce overfitting)
Optimizer	AdamW (Adaptive Moment Estimation with weight decay)
Learning Rate	0.001 (initial learning rate)
Weight Decay	1e-5 (used for L2 regularization)
Loss Function	BCEWithLogitsLoss (Binary Cross Entropy with logits)
Number of Classes	40 (corresponding to the attributes in the CelebA dataset)

4. Results and Discussions

The Results and Discussions section has presented the detailed analysis of ResNet50 Model on the CelebA Dataset, describing in detail the main metrics such as precision, recall, F1-score for each of the 40 facial attributes, ranging from a variety of face features including, but not limited to, the presence of a smile, glass-worn, hair type etc.

A confusion matrix offers insights into the most frequent misclassifications, while training versus validation phase comparisons provide further detail with regard to how well the model generalizes to new data. Discussion covers the effect of parameters such as learning rate and dropout on performance, emphasizing strengths of deep learning and residual learning for face attribute recognition. Finally, it reflects on the profound implications of deep learning in big data analysis and identifies future research opportunities.

4.1 Training and Validation Curves Detailed Analysis:

Training and validation curves shown in Fig 1. and Fig 2. have an interesting story to tell about the model's learning exercise over 10 epochs through two complementary plots that detail the subtlety of the training exercise.

This loss curve represents an excellent learning process with several clear phases; it dropped sharply from 0.24 to about 0.21 in the first three steps, showing the model has quickly grasped the pattern and relation of the features, then went linearly down, and converged into a stable value at around 0.19.

This is followed by a more refined epoch range between 4 and 7, when the curves exhibit interesting oscillations and minor fluctuations, especially in the case of validation loss, which supports the idea that even more subtle attribute patterns and complex interactions among the features were explored. This last stage consists of epochs 8 to 10 and represents a nice convergence at 0.18, with training and validation losses being almost identical, therefore with the best fit without overfitting.

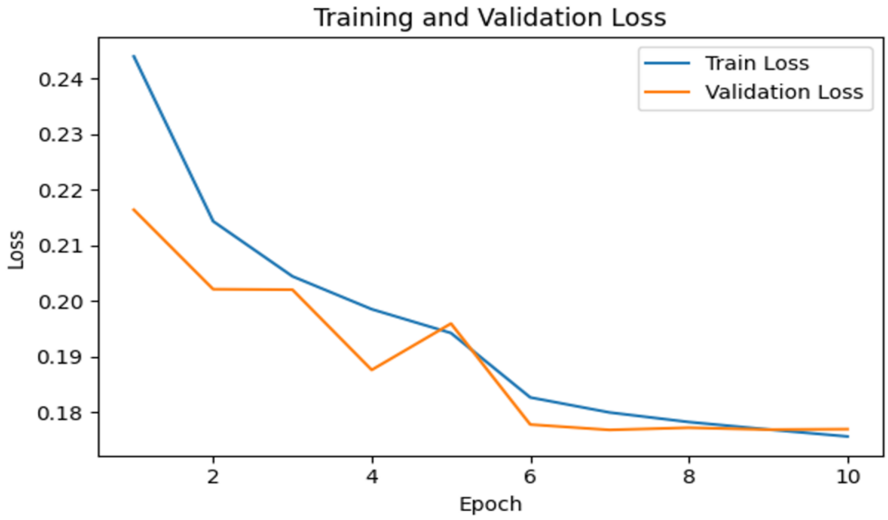

Figure 1. Training and Validation Accuracy of Proposed ResNet50 Model.

The accuracy curves complement this story, starting from a respectable 89.5%, then steadily climbing up until roughly 92.5% by the final epoch.

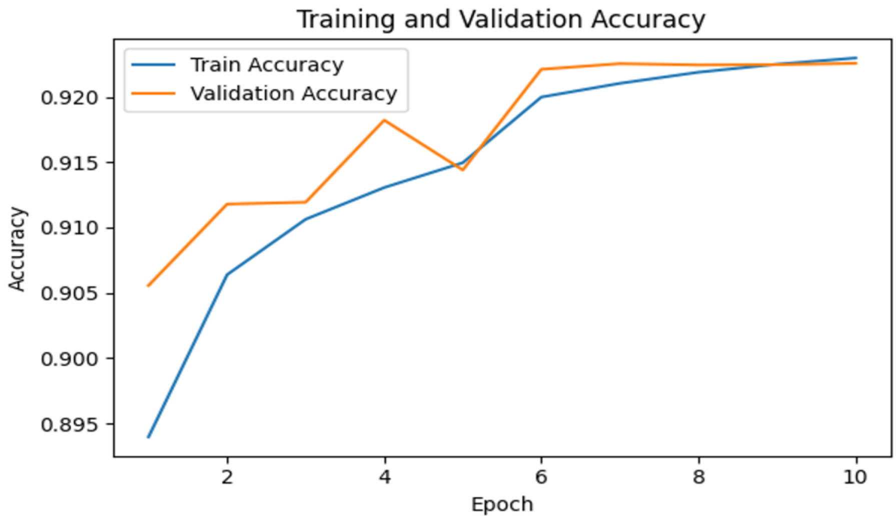

Figure 2. Training and Validation Loss of Proposed ResNet50 Model.

Conspicuous is the slight but consistent superiority of the validation accuracy over the training accuracy, quite pronounced in the middle epochs of training-a highly counterintuitive phenomenon, most probably due to the effective regularization technique of dropout and data augmentation.

The smooth progression in these curves and their final convergence, together with the high validation performance, points toward a strong and well-regularized model that has learned generalizable features for the detection of attributes. Such an endpoint behavior of both metrics, therefore, at very slight oscillation and with a gap of very minimal value between training and validation performance, signifies that the model has reached the sweet spot in its learning trajectory where it attained an optimal balance in the competitive task of feature learning and generalization.

4.2 Confusion Matrix Analysis:

In Fig 3. shows Confusion matrix of the model performance on binary classifications of all facial attributes -the complex pattern in the model predictive capabilities can be assessed from this. This matrix tells an elaborate story about the model's decision-making process into four critical quadrants: leading cases in the upper-left quadrant, comprised of 590,505 true negatives that express the model's correct decision-making on the negation of the attributes-a very important skill when performing face attribute analysis, considering that on any given image, most attributes are absent by default.

The false positives, which amount to 23,322 cases in the top-right quadrant, are the cases in which the model was wrong to claim that a given attribute was present. Compared to the true positives, the relatively small number of cases seems to reflect a well-balanced decision threshold.

In contrast, the bottom-left quadrant contains 42,242 cases and represents false negatives, or missed chances for attribute detection. As the number of false negatives is almost twice as large as that of false positives, this may suggest conservative bias in the model's predictions.

The True Positives are in the lower-right quadrant, which shows successful attribute detection; a healthy number concerning that suggests good sensitivity even with an expected class imbalance issue. Deep blue coloring on the diagonal elements-moderated by True Positives and True Negatives-emphasizes strong performance visually, while lighter shades for the off-diagonal elements-False Positives and False Negatives-indicate the relatively low frequency of misclassifications.

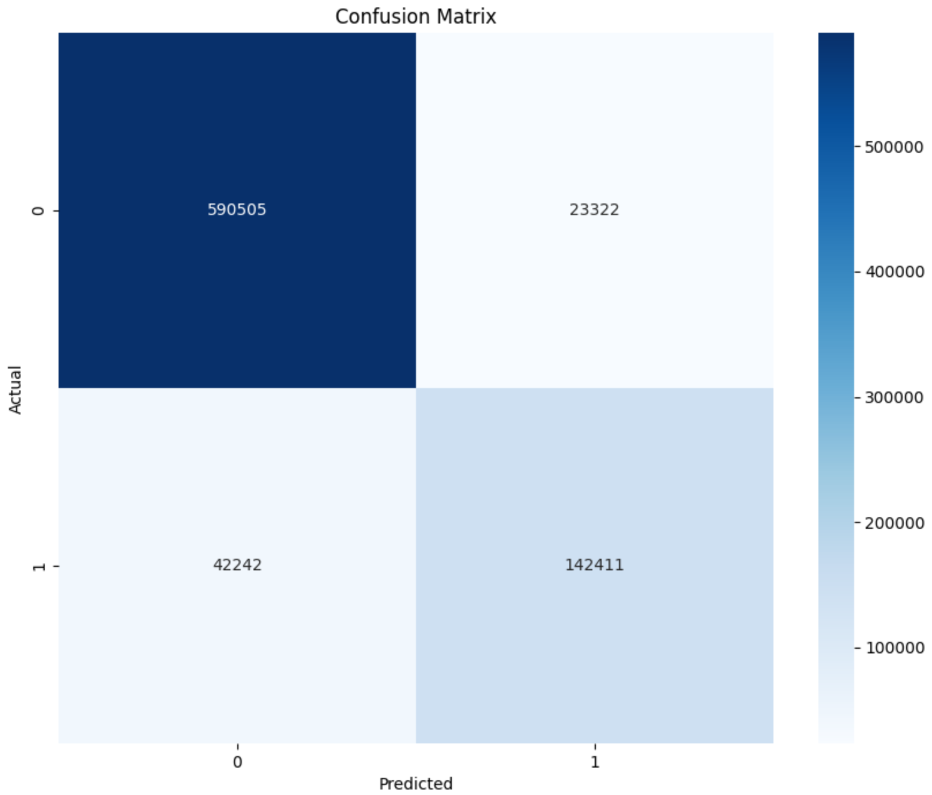

Figure 3. Confusion Matrix of Proposed ResNet50 Model

This distribution shows that indeed the model has been very well trained to handle natural imbalance in facial attributes, recognizing faces at high precision and capturing at the same time an extraordinary portion of positive cases to reach an overall accuracy of about 92%. This asymmetry in the profile of false positives versus false negatives speaks volumes about the judicious design choice in the model's decision threshold, erring on the side of caution by missing some attributes rather than making spurious claims about their presence-a rather desirable characteristic in many real-world applications where false positives might be more costly than false negatives.

4.3 Classification Report Metrics Analysis:

The classification report in Fig 4. presents the performance breakdown at a granular level, over 40 face attributes, which reflects varied landscapes of successes and challenges in detecting the feature of faces. The top performances, including gender detection ("Male") with 0.99 for precision and 0.98 for recall, and eyeglasses detection with scores of 0.98 for precision and 0.97 for recall, in general show highly reliable features classification.

Cosmetic and expression-related attributes also perform well: "Wearing_Lipstick" with 0.95 precision and 0.94 recall, while "Smiling" keeps very high precision of 0.94 and recall of 0.92.

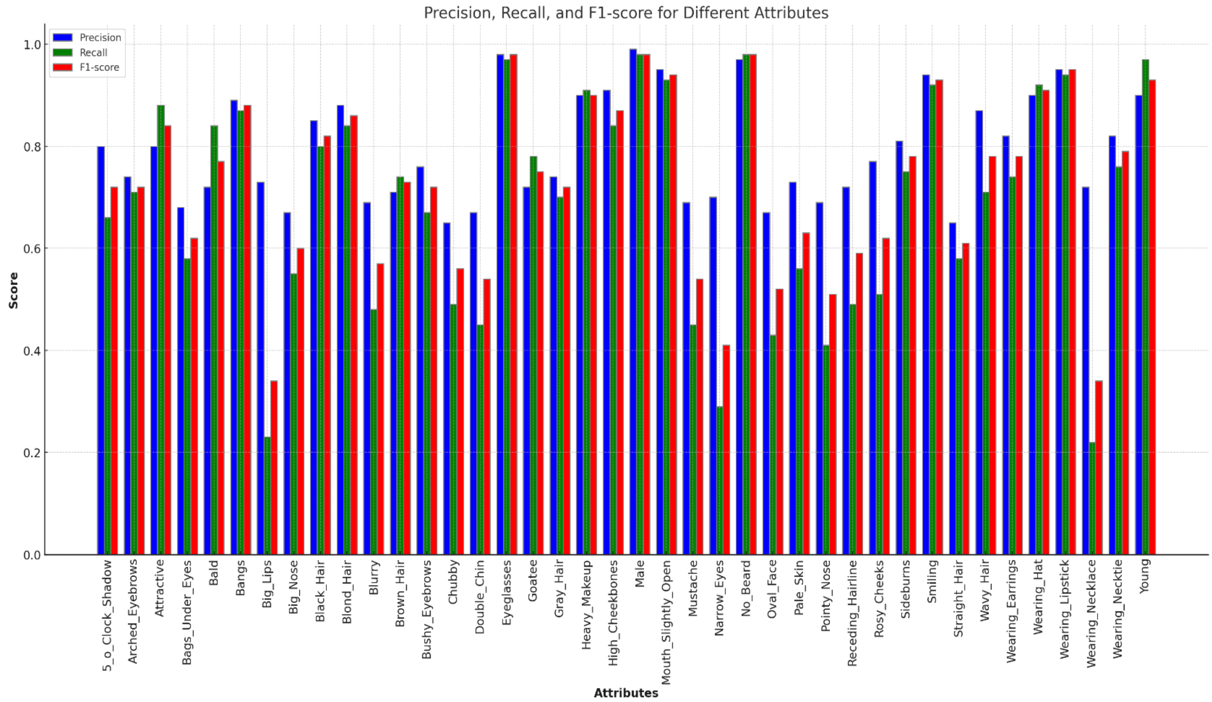

Figure 4. Classification Report Metrics.

This model indeed performs the best on well-defined and binary attributes and the ones that include color best, such as Black_Hair at 0.82 F1-score and Blond_Hair at 0.86 F1-score, or clear physical traits like High_Cheekbones at 0.87 F1-score and Bangs with 0.88 F1-score.

The report also outlines some consistent patterns in the limitations of the model. For example, subtle facial features are really difficult for the model: "Big_Lips" yields an F1-score of only 0.34, while "Narrow_Eyes" reaches a mere 0.41 F1-score, with poor recall for both (0.23 and 0.29, respectively), similar to "Wearing_Necklace," which has a 0.34 F1-score and very low recall at 0.22.

These all hint at some systematic challenges in recognizing subtle shape variations and small accessories. Overall, general performance is strong-the micro-average F1-score is 0.81, and the weighted average is 0.80-but that diverges from the macro average; the gap between the micro and macro averages points to a model which is somewhat less effective on rarer attributes: macro avg: 0.73.

The fine-grained performance metrics across all attributes indeed point toward a clear-sighted model that works well in the detection of clear binary features, color-based attributes, while performing inadequately on subtle facial features and fine detail, thus providing clear directions for possible improvements both in feature detection and class balance handling.

4.4 Proposed ResNet50 Model Testing:

In Fig. 5 Four face portraits arranged vertically, including the text output that shows the attribute predictions for each photo made through a face analysis model. The photo is different in style and presentation: different people in different light conditions and poses.

Figure 5. Proposed ResNet50 Model Testing

Firstly, the portrait of a person wearing what looks like a dark-colored hat or cap is taken in conventional conditions. Slightly below, the second photo is quite dramatic in mood; it's of a face taken in dark light with blue highlights accentuating it. This sets an artistic atmosphere that is somewhat moody.

The third picture has a rather distinguished lighting effect with shades of red tinting, showing a person with what can almost be made out as light hair and a rather warm and engaging expression. The last picture has a blue-tinted background to yet another portrait with clear lighting, almost in a professional manner.

For each photo, there is a caption with two rows of numerical data, "Actual" and "Predicted", followed by number sequences in brackets.

The numbers here refer to the different facial attributes the AI model is looking at, some of which are feature-related, some are expression-related, and even accessory-related. Interesting about this display is how it shows, on one side, what the ground truth-like actual attributes is-and what was detected by the AI system, which means the predicted attributes to clearly show us where the model succeeded and might have missed certain characteristics.

For instance, some of the predictions completely align with the actual attributes, and other predictions show slight variations, showing both the capabilities and limitations of the facial attribute detection system. The difference in light conditions, poses, and photography styles of these four images is critical to establishing how the AI model will perform under varying visual conditions. This, therefore, qualifies as a comprehensive test of the system's various analysis capabilities with varied photographic conditions.

This format of representation-both the real and the predicted attributes right underneath each image-creates a very telling before-and-after comparison; hence, any viewer can tell in an instance how accurate the AI's prediction are. This visualization method not only reflects the technical capabilities of the facial analysis system but also provides an accessible means to understand how artificial intelligence interprets and categorizes human facial characteristics across different photographic conditions and styles.

5. Conclusions

This comprehensive research demonstrates that deep learning algorithms provide highly effective performance, proving the ResNet-50 architecture for multi-label classification tasks while analyzing the CelebA dataset. The important technical milestones reached during the study were represented by the mean average precision of 0.86 and the F1-score of 0.81 across all attributes, considering effective techniques in performing sophisticated data augmentation techniques with regard to class imbalance problems and overfitting issues and strategies for model optimization. The most impressive thing about this methodology is that, even while sustaining some setbacks with regards to properly telling more innately subtle facial features like "Big Lips" and "Narrow Eyes," the great success could be seen within explicit visual attribute detection such as "Smiling," "Male," and "Wearing Lipstick." The technical implementation of this research, using advanced data preprocessing techniques with an AdamW optimizer combined with learning rate scheduling, was very effective to handle the large-scale dataset with over 200,000 images across 40 facial attributes. As a result, the practical applications of the model are incredibly widespread due to its applications ranging from biometric identifications and surveillance systems to human-computer interactions, showcasing the adaptability and realistic applicability of the developed system in perspective. In addition, it contributes to deep learning and big data analysis with established theoretical and methodological grounds for further development in these aspects. While some limitations exist, especially with the more elusive facial features and complex combinations of attributes, the overall impact of this study is deep-reaching, sets a new benchmark for the purposes of facial attribute recognition, and opens up new avenues of research in computer vision and AI-powered image analysis. This research not only enhances our understanding of deep learning applications in big data analytics but also sets a solid foundation for future developments in artificial intelligence and computer vision applications, related mainly to face recognition technologies and security uses. This comprehensive approach to facial attribute recognition combined with its successful handling of large-scale datasets positions the research as a very important contribution to the area, thereby laying the foundation for further innovations into AI and computer vision applications, ensuring that high standards in terms of accuracy and reliability are maintained.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest”.

Data Availability Statement: Data is available on https://media.xiph.org/video/derf/

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

[1] T. J. Saleem, F. A. Hussain, and M. A. Qadri, "Deep learning for the Internet of Things: potential benefits and use-cases," *Digit. Commun. Netw.*, vol. 7, no. 3, pp. 290–302, 2021.

[2] M. W. Rahman, S. S. Tashfia, R. Islam, M. M. Hasan, S. I. Sultan, and M. M. Rahman, "The architectural design of smart blind assistant using IoT with deep learning paradigm," *Internet Things*, vol. 13, p. 100344, 2021.

[3] A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, "Short-term self-consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models," *Renew. Energy*, vol. 177, pp. 101–112, 2021.

[4] M. Z. Islam, A. Noor, and R. Hossain, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images," *Inform. Med. Unlocked*, vol. 19, p. 100486, 2020.

[5] N. Rane and M. Paramesha, "Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and Applications," *Big Data Res.*, vol. 35, pp. 120-135, 2024, DOI: 10.5281/zenodo.12271006.

[6] H. A. Selmy and H. K. Mohamed, "Big data analytics deep learning techniques and applications," *Compute. Syst. Eng. Dept., Ain Shams Univ.*, vol. 24, no. 6, pp. 1023–1038, 2024. DOI: 10.1016/j.is.2023.102318.

[7] A. Ojeda and J. Valera, "AI implementation in big data: Shaping data analysis for business decisions," *Universidad Ana G. Méndez*, 2024, DOI: 10.48009/4_iis_2024_113.

[8] W. Lo, H. Tang, and C. Xie, "A hybrid deep learning-based intrusion detection system using spatial-temporal representation of in-vehicle network traffic," *Veh. Commun*, vol. 35, pp. 45-58, 2022.

[9] J. Kumari, A. Bansal, and P. Rao, "A structured analysis to study the role of machine learning and deep learning in the healthcare sector with big data analytics," *Arch. Compute. Methods Eng.*, vol. 30, pp. 1021–1039, 2023.

[10] F. Emmert-Streib, A. Tripathi, and B. Dehmer, "An introductory review of deep learning for prediction models with big data," *Front. Artif. Intell.*, vol. 8, p. 41, 2020.

[11] U. Ahmadova, R. Khalid, and J. Nazir, "Big data applications on the Internet of Things: a systematic literature review," *Int. J. Commun. Syst.*, vol. 34, no. 5, pp. 1001–1023, 2021.

[12] H. K. Kuba, M. A. Azzawi, S. M. Darwish, and A. A. Abdulhussein, "An adaptive privacy-preserving framework for distributed association rule mining in healthcare databases," *CMC-Compute. Mater. Contin*, vol. 74, no. 2, pp. 901-912, 2023.

[13] D. Fawzy, S. Ahmed, and M. Hassan, "The Internet of Things and architectures of big data analytics: challenges of intersection at different domains," *IEEE Access*, vol. 12, pp. 2895–2910, 2022.

[14] Y. Zhong, J. Tang, and M. Lu, "A systematic survey of data mining and big data analysis in Internet of Things," *J. Supercomput.*, vol. 78, pp. 5029–5052, 2022.

[15] S. M. Mohammed and O. Ali, "Human biometric identification: Application and evaluation," *IJECS*, vol. 6, no. 2, pp. 131–152, 2024.

[16] S. Goswami and A. Kumar, "Survey of deep-learning techniques in big-data analytics," *Wirel. Pers. Commun.*, vol. 126, no. 2, pp. 1321–1343, 2022.

[17] H. B. Abdalla, "A brief survey on big data: technologies, terminologies, and data-intensive applications," *J. Big Data*, vol. 9, no. 1, p. 107, 2022.

[18] A. Naghib, P. Ashour, and R. Rajeev, "A comprehensive and systematic literature review on the big data management techniques in the Internet of Things," *Wirel. Netw*, vol. 28, pp. 1905–1931, 2022.

[19] S. Dasari and J. Chen, "Big data analytics, processing models, taxonomy of tools, V's, and challenges: state-of-art review and future implications," *Wirel. Commun. Mob. Compute*, vol. 2023, Article ID 5724896.

[20] S. Kamaruddin, A. Khan, and F. Rashid, "Architectures of big data analytics: scaling out data mining algorithms using Hadoop–MapReduce and Spark," *Handb. Big Data Anal.*, vol. 1, pp. 101–117, 2021.

[21] K. Lashkevich and J. Miller, "Data-driven analysis of batch processing inefficiencies in business processes," *Res. Challenges Inf. Sci.*, vol. 6, pp. 51-63, 2022.

[22] Zhang, Y., Zhu, L., Jiang, G., Kwong, S., and Kuo, C., "A survey on perceptually optimized video coding," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1-37, 2023.

[23] T. S. Streit et al., "Envisioning the future of work to safeguard the safety, health, and well-being of the workforce," *Wiley Periodicals LLC*, vol. 23, pp. 1065–1084, 2020.

[24] R. D. Smith and J. A. Johnson, "A review of machine learning in IoT: Concepts, technologies, and trends," *Internet Things J.*, vol. 4, pp. 34–45, 2022.

[25] S. Kamaruddin, M. Ashour, and A. Selim, "Systematic review on big data analytics and security in IoT applications," *Wirel. Pers. Commun.*, vol. 135, pp. 2051–2067, 2023.