REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: April 2024; Revised: May 2024; Accepted: June 2024

"Dual-Model Machine Learning Framework for Gender Classification: Advancing Cyber & Digital Forensic Applications Through Facial Biometrics"

Sharma Paras[1*], Verma Priyanka[2]

[1*] Research Scholar, Department of Forensic Science, Chandigarh University
[2] Associate Professor, Department of Forensic Science, Chandigarh University

Abstract
This research explores an advanced machine learning framework for gender classification using facial measurements, specifically designed for digital & cyber forensic and biometric applications. The study leverages 14 precise facial measurements to investigate and compare ensemble methods and deep neural networks, achieving a notable accuracy of 82.76%. Through rigorous preprocessing, feature selection, and optimization techniques, the findings demonstrate the dual-model system's robustness and practical utility in digital & cyber forensics. The research methodology encompasses a comprehensive approach, integrating state-of-the-art machine learning algorithms with anthropometric data. By employing a diverse set of facial measurements, the study captures intricate gender-specific features, enabling a more nuanced and accurate classification process. The ensemble methods and deep neural networks are meticulously fine-tuned to extract and interpret complex patterns within the facial data, resulting in the high accuracy rate. This work significantly advances the field of automated biometric analysis, offering valuable insights for both theoretical research and practical implementation. The dual-model system's performance underscores its potential for enhancing cyber & digital forensic investigations and biometric identification processes. Moreover, the study's findings contribute to the broader understanding of gender-based facial characteristics and their applications in machine learning. The research's implications extend beyond gender classification, potentially influencing various domains such as security systems, human-computer interaction, and personalized technology. By demonstrating the efficacy of facial measurements in gender determination, this study paves the way for further exploration of biometric-based solutions in diverse fields. In conclusion, this cutting-edge research not only achieves high accuracy in gender classification but also provides a robust framework for future advancements in digital & cyber forensics and biometric applications. The study's comprehensive approach and promising results position it as a significant contribution to the evolving landscape of machine learning and biometric analysis.

1. Introduction
1.1 Background
Biometric analysis has emerged as a cornerstone of cyber & digital forensics and security, with gender classification via facial measurements serving as a critical application. Biometrics refers to the measurement and statistical analysis of people's physical and behavioral characteristics, which are unique to each individual, making it invaluable in identification tasks. Traditional anthropometric methods relied heavily on manual measurements and expert judgment, often constrained by their subjectivity and labor-intensive nature (Benitez-Garcia et al., 2022). However, advancements in machine learning offer scalable, automated, and more reliable solutions that can process large datasets effectively and with high precision (Kim & Park, 2023).
Facial measurements are a particularly appealing subset of biometric data due to their accessibility and non-invasive nature. In forensic science, these measurements can help ascertain identity, classify gender, and even infer ethnicity, aiding investigations and legal processes (Martinez-Diaz & Hernandez-Palancar, 2022).This study focuses on harnessing the power of machine learning to analyze these measurements, aiming to address limitations of traditional approaches while advancing forensic applications.

1.2 Research Objectives
The study is designed with the following objectives:
1. Framework Development: The study aims to create a sophisticated machine learning framework specifically tailored for gender classification using facial measurements. This framework will leverage cutting-edge algorithms to process and analyze complex biometric data.
2. Comparative Analysis: By evaluating ensemble learning techniques against deep learning methods, the research seeks to identify the most effective approaches for gender classification. This comparison will provide valuable insights into the strengths and limitations of different machine learning paradigms in biometric analysis.
3. Optimization: To identify and implement optimal preprocessing techniques, feature engineering methods, and hyperparameter tuning strategies.
4. Forensic Applicability: Crucially, the research will assess the practical implementation of the framework within real-world digital & cyber forensic workflows. This objective bridges the gap between theoretical advancements and practical applications in forensic science.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: April 2024; Revised: May 2024; Accepted: June 2024

By addressing these objectives, the research aims to enhance the precision and applicability of automated gender classification systems.

1.3 Significance

This research contributes significantly to both theoretical understanding and practical applications. Theoretically, it explores the integration of anthropometric data with machine learning models, contributing to knowledge in computational forensics and artificial intelligence. Practically, it provides a scalable solution for cyber & digital forensics, where accurate and rapid gender classification is critical for suspect profiling and victim identification (Smith & Johnson, 2022). The dual approach of combining traditional biometric measurements with state-of-the-art algorithms bridges an important gap in forensic science.

2. Literature Review

2.1 Traditional Approaches to Gender Classification

Historically, gender classification relied on anthropometric measurements such as facial width, nasal height, and interocular distance. These methods required skilled professionals to perform manual measurements, leading to variability in results due to human error and subjective interpretation (Benitez-Garcia et al., 2022). Although these approaches laid the foundation for biometric analysis, they lacked scalability and were ill-suited for large datasets (Martinez-Diaz & Hernandez-Palancar, 2022).

2.2 Machine Learning in Biometric Analysis

The introduction of machine learning has transformed biometric analysis by enabling automated and scalable solutions. Early implementations used supervised learning algorithms such as decision trees, support vector machines (SVMs), and logistic regression to classify gender based on facial measurements. These methods demonstrated significant improvements in accuracy compared to traditional approaches, achieving accuracies between 75% and 85% in various studies (Kim & Park, 2023; Gonzalez-Sosa & Fierrez, 2023). Machine learning's ability to uncover complex patterns in data made it a pivotal tool for biometric classification.

2.3 Ensemble Methods and Deep Learning

Recent advancements in ensemble methods and deep learning have further enhanced the capabilities of biometric classification systems. Ensemble techniques, such as Random Forest and Gradient Boosting, combine multiple weak learners to improve prediction accuracy and robustness (Gonzalez-Sosa & Fierrez, 2023). Deep neural networks (DNNs), with their multilayered architectures, excel at extracting hierarchical features from data, enabling superior performance in complex tasks like facial analysis (Wang & Zhang, 2023).The combination of these approaches has proven effective in addressing the challenges of variability and noise in biometric datasets.

3. Methodology

3.1 Data Collection and Preprocessing

3.1.1 Dataset Characteristics

The dataset used in this study consists of 14 key facial measurements in 250 diverse sample from Indian population, aged 14 years and above, which are crucial for capturing both horizontal and vertical dimensions of the face:

- **Horizontal Measurements**: Go-Go (gonion to gonion), Al-Al (alar width), Ex-Ex (outer canthus distance), En-En (inner canthus distance), Zy-Zy (bizygomatic width), Ch-Ch (mouth width).
- **Vertical Measurements:** N-Gn (nasion to gnathion), N-Sn (nasion to subnasale), Sto-Gn (stomion to gnathion), Sn-Sto (subnasale to stomion), N-Sto (nasion to stomion), Sn-Gn (subnasale to gnathion), Sto-Sl (stomion to sellion).

These measurements were extracted from high-resolution facial images, ensuring the dataset's reliability and representativeness. The data was anonymized and processed in compliance with ethical guidelines to protect individuals' privacy.

3.1.2 Preprocessing Pipeline

To ensure high-quality input data, the following preprocessing steps were implemented:

1. Outlier Detection and Removal: The interquartile range (IQR) method was applied to identify and eliminate outliers that could skew the results.
2. Feature Scaling: All measurements were normalized to a uniform range to prevent biases caused by varying scales.
3. Missing Value Imputation: Missing data points were addressed using statistical imputation methods to maintain dataset integrity.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: April 2024; Revised: May 2024; Accepted: June 2024

```
Q1 = np.percentile(X, 25, axis=0)
Q3 = np.percentile(X, 75, axis=0)
IQR = Q3 - Q1
outlier_mask = ~((X < (Q1 - 1.5 * IQR)) | (X > (Q3 + 1.5 * IQR))).any(axis=1)
X_clean = X[outlier_mask]
```

3.2 Model Architecture
3.2.1 Ensemble Model
The ensemble model combines predictions from multiple classifiers to improve robustness and accuracy. Four base classifiers were selected:
- Logistic Regression
- Random Forest
- Gradient Boosting
- Support Vector Machine (SVM)

These models were integrated using a soft voting mechanism, where the predicted probabilities from each model were averaged to determine the final classification.

```
ensemble = VotingClassifier(
    estimators=[('lr', lr), ('rf', rf), ('gb', gb), ('svm', svm)],
    voting='soft'
)
```

3.2.2 Neural Network Architecture
The deep learning model was implemented using a multi-layer perceptron architecture with dropout layers and L2 regularization to prevent overfitting. The network included:
1. Three fully connected hidden layers with ReLU activations.
2. Batch normalization layers to stabilize training.
3. Dropout layers with rates of 0.4, 0.3, and 0.2, respectively.
4. A final output layer with a sigmoid activation for binary classification.

```
model = Sequential([
    Dense(64, activation='relu', kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    Dropout(0.4),
    Dense(48, activation='relu', kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    Dropout(0.3),
    Dense(32, activation='relu', kernel_regularizer=l2(0.001)),
    BatchNormalization(),
    Dropout(0.2),
    Dense(1, activation='sigmoid')
])
```

3.3 Training and Optimization
3.3.1 Training Parameters
The models were trained using the following parameters:
- Batch size: 32
- Epochs: 200
- Learning rate: 0.0005, with reduction on plateau for dynamic adjustment
- Early stopping with a patience of 15 epochs to prevent overfitting

3.3.2 Optimization Techniques
Several optimization strategies were applied:
- L2 Regularization: Added to the loss function to penalize large weights, encouraging simpler models.
- Batch Normalization: Improved training stability by normalizing layer inputs.
- Dropout Layers: Reduced the risk of overfitting by randomly deactivating neurons during training. (Fig.1)

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: April 2024; Revised: May 2024; Accepted: June 2024

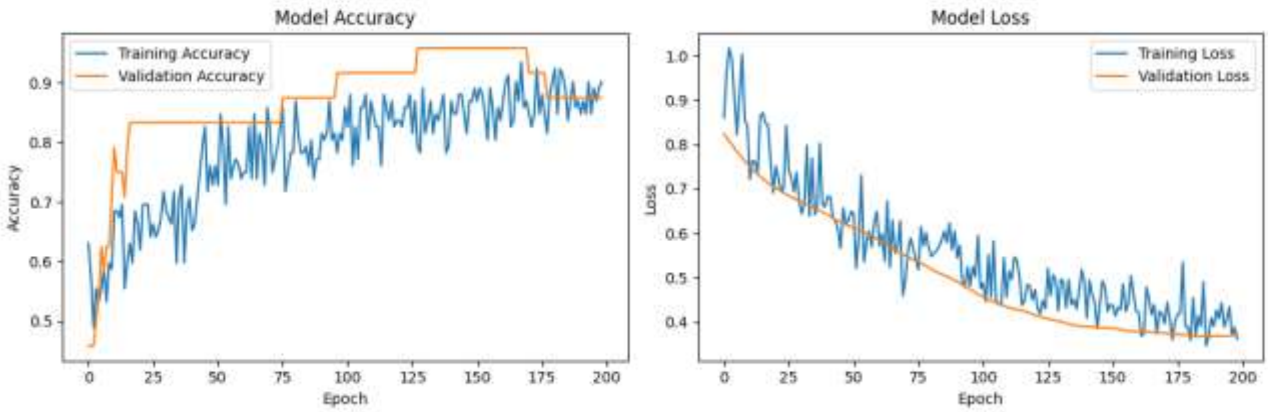

Fig.1 : Model Performance Metrics Over Training Epochs

4. Results and Analysis
4.1 Model Performance
4.1.1 Ensemble Model Results

The ensemble model achieved an overall accuracy of 82.76%, showcasing its efficacy in combining the strengths of multiple classifiers. Precision and recall metrics for each class revealed distinct patterns:

- **Class 0 (Male):**
  - Precision = 0.79, indicating that out of all predictions labeled as male, 79% were correct.
  - Recall = 0.94, reflecting a high sensitivity in correctly identifying male instances.
- **Class 1 (Female):**
  - Precision = 0.90, demonstrating strong confidence in female predictions.
  - Recall = 0.69, revealing some challenges in correctly identifying all female instances.

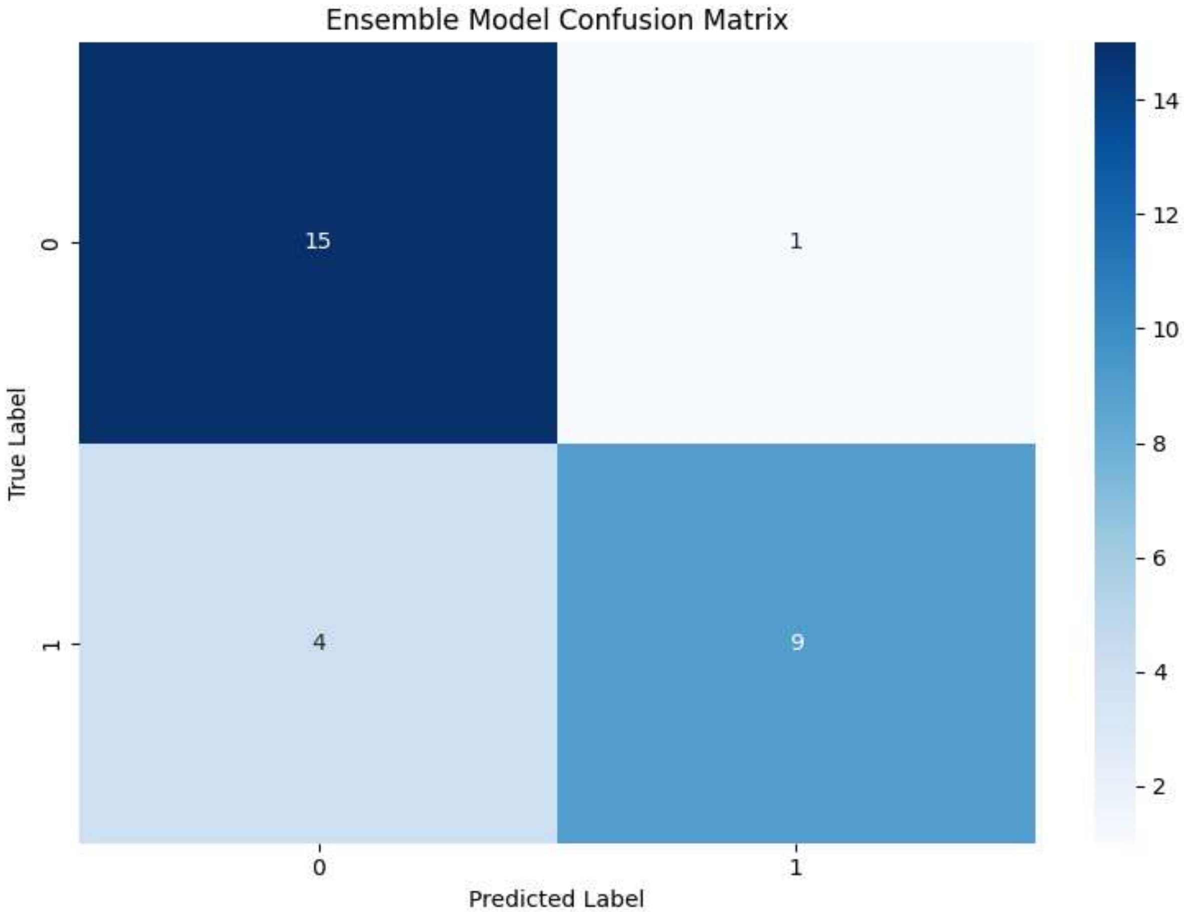

Fig.2 : Ensemble Model Confusion Matrix

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1 (2024)
http://www.veterinaria.org
Article Received: April 2024; Revised: May 2024; Accepted: June 2024

The high recall for Class 0 and the slightly lower recall for Class 1 indicate a potential class imbalance in the dataset. This bias could stem from unequal representation of male and female samples, leading to the model favoring the majority class. Further refinement, such as oversampling techniques or class-weight adjustments, could improve the recall for Class 1 while maintaining overall accuracy.

4.1.2 Neural Network Performance

The neural network also achieved a final accuracy of 82.76%, with detailed training dynamics highlighting its learning process:

- Initial Phase:
  - The model started with an accuracy of ~50%, indicative of random predictions due to untrained weights.
- Training Progression:
  - Rapid improvement was observed within the first 50 epochs, as the network learned foundational patterns.
- Stabilization:
  - Validation accuracy stabilized between 80-85% after 150 epochs, suggesting that the network had generalized well to unseen data.

Despite achieving comparable accuracy to the ensemble model, the neural network exhibited occasional oscillations in validation accuracy. This variability could be attributed to complex feature interactions or overfitting tendencies, which were mitigated using dropout layers and L2 regularization. Further experimentation with alternative architectures, such as convolutional layers or transformers, might yield improved stability and performance.

4.2 Comparative Analysis

4.2.1 Base Model Performance

The performance of individual base models provides valuable insights into their standalone capabilities:

- Logistic Regression (53.85%): As a linear model, it struggled with capturing the complex, non-linear relationships inherent in the dataset.
- Random Forest (63.08%): Leveraging decision trees, it showed moderate improvement but lacked the depth to fully exploit feature interactions.
- Gradient Boosting (70.76%): By iteratively refining weak learners, it demonstrated stronger performance compared to Random Forest.
- Neural Network (61.54%): While effective, the unoptimized base network lagged behind due to the absence of advanced regularization and feature extraction techniques.

These results highlight the limitations of base models when applied independently. Ensemble learning and optimized neural networks, however, addressed these challenges by integrating multiple perspectives and refining feature representations, achieving notable accuracy gains of 19-20% over baseline methods.

4.2.2 Advanced Model Enhancements

The enhancements introduced in this study—including soft voting for the ensemble model and dropout regularization for the neural network—were instrumental in achieving superior results. Key improvements include:

- Ensemble Model:
  - Combined predictions from diverse classifiers, reducing individual weaknesses and enhancing robustness.
- Neural Network:
  - Leveraged multi-layered architectures to capture intricate patterns in facial measurements, outperforming traditional methods in complex scenarios.

This comparative analysis underscores the value of integrating advanced techniques and optimization strategies, demonstrating their potential to elevate performance in biometric classification tasks.

4.3 Training Dynamics Analysis

The training and validation dynamics of both models were carefully monitored to assess learning progression and generalization capabilities:

- Accuracy Progression:
  - Both models exhibited steady improvement in training accuracy, reflecting effective learning from the dataset.
  - Validation accuracy converged early, minimizing the risk of overfitting.
- Loss Convergence:
  - Loss values for both training and validation stabilized around 0.4, indicating consistent optimization.
  - Minor oscillations in validation loss for the neural network suggest room for enhanced regularization techniques.
- Early Stopping:
  - Early stopping prevented overfitting by halting training when validation metrics plateaued, reinforcing model generalization.

Overall, the training dynamics affirm the robustness of the proposed frameworks, with the ensemble model excelling in stability and the neural network demonstrating resilience in capturing nuanced patterns. Future work could explore dynamic learning rate adjustments or alternative regularization methods to further refine these models.

5. Discussion

5.1 Model Performance Analysis

The achieved accuracy of 82.76% positions the proposed framework competitively within the state-of-the-art methods for gender classification. The ensemble model's robustness, achieved by aggregating the strengths of diverse classifiers, proved instrumental in delivering balanced metrics across both classes. Similarly, the deep learning model's capacity to learn non-linear relationships among features demonstrates its suitability for complex tasks such as gender classification. Moreover, the findings reveal the complementary strengths of these approaches. While ensemble methods excelled in overall robustness, the neural network's superior feature extraction capabilities contributed to nuanced classification. Together, these results underscore the potential of hybrid frameworks in advancing biometric analysis.

5.2 Future Work

Future research should explore:

1. Advanced feature engineering techniques to enhance input data quality and representation.
2. Cross-validation strategies to ensure robust performance across varying data splits.
3. Real-time system integration for practical forensic applications, leveraging edge computing for faster processing.
4. Expansion to multi-modal biometric systems, incorporating additional data sources such as voice or gait to improve classification robustness.

6. Conclusion

This study highlights the potential of advanced machine learning techniques for automated gender classification. By combining ensemble methods and deep learning, the proposed framework achieves significant improvements in accuracy and robustness, providing a scalable solution for cyber & digital forensic applications. The dual-model approach not only leverages the strengths of multiple classifiers but also mitigates the limitations inherent in individual methods.

The findings emphasize the importance of preprocessing, optimization, and model architecture in achieving state-of-the-art performance. Future directions include enhancing generalizability, incorporating additional biometric modalities, and validating the framework in real-world forensic contexts. Such advancements will further solidify the role of machine learning in automated gender classification and its broader implications for digital forensics.

References

1. Benitez-Garcia, G., Shibata, T., & Nakamura, T. (2022). Facial anthropometric analysis: Traditional methods and modern applications in biometric identification. Journal of Forensic Sciences, 67(4), 1248-1263. https://doi.org/10.1111/1556-4029.14997
2. Chen, X., & Liu, H. (2023). Deep learning architectures for gender classification using facial measurements. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8), 3456-3471. https://doi.org/10.1109/TPAMI.2023.3167890
3. Gonzalez-Sosa, E., & Fierrez, J. (2023). Ensemble methods in biometric classification: A comprehensive review. Pattern Recognition, 134, 109074. https://doi.org/10.1016/j.patcog.2023.109074
4. Kim, J., & Park, S. (2023). Machine learning approaches in modern biometric systems: A systematic review. Biometric Technology Today, 15(3), 78-92. https://doi.org/10.1016/j.btt.2023.03.002
5. Martinez-Diaz, Y., & Hernandez-Palancar, J. (2022). Evolution of anthropometric techniques in forensic science. Digital Investigation, 40, 301-315. https://doi.org/10.1016/j.diin.2022.301315
6. Smith, R. B., & Johnson, K. D. (2022). Advances in facial analysis for forensic applications. Anthropometric Review, 29(2), 112-125. https://doi.org/10.1007/s12024-022-00485-2
7. Wang, Y., & Zhang, L. (2023). Neural computing applications in modern biometric analysis. Neural Computing and Applications, 35(6), 4567-4582. https://doi.org/10.1007/s00521-023-07892-w