



Original Research Paper

# GCAN: A Gender-Conditioned Attentive Network with Distribution-Aware Learning for Joint Age and Gender Estimation from Facial Images

S Vimal<sup>1</sup>, M. Kaliappan<sup>2</sup>, Gaurav Dhiman<sup>3,4,5</sup>

<sup>1</sup>Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India

<sup>2</sup>Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India

<sup>3</sup>Department of Computer Science and Engineering, Yuan Ze University, Tao Yuan, Taiwan

<sup>4</sup>University Centre for Research and Development, Chandigarh University, Mohali, India

<sup>5</sup>Division of Research and Development, Lovely Professional University, Phagwara, India

Email: [kaliappan@ritrjpm.ac.in](mailto:kaliappan@ritrjpm.ac.in), [svimal@kitcbe.ac.in](mailto:svimal@kitcbe.ac.in), [gdhiman0001@gmail.com](mailto:gdhiman0001@gmail.com)

**Abstract**—Automated estimation of demographic attributes such as age and gender from facial images is increasingly relevant to health informatics, supporting patient triage in telemedicine, demographic auto-population of electronic health records, cohort stratification, and forensic and prognostic applications. Although these two attributes are physiologically coupled facial ageing manifests differently across sexes—most deep learning systems predict them with independent output heads and combine their losses using hand-tuned static weights, discarding this coupling and introducing brittle hyperparameters. This paper proposes GCAN, a Gender-Conditioned Attentive Network built on a Swin Transformer backbone that addresses these limitations through three coordinated mechanisms. First, a Feature-wise Linear Modulation (FiLM) block allows gender-aware features to generate affine parameters that directly modulate the age representation, realising an explicit directed task coupling. Second, age is predicted as a soft distribution over discrete age bins and optimised with a mean-variance loss, yielding a robust expectation-based estimate rather than a brittle point regression. Third, the age and gender losses are balanced automatically through homoscedastic uncertainty weighting, eliminating manual loss-weight tuning. On the UTKFace benchmark, GCAN attains a representative gender-classification accuracy of **95.8%** and an age mean-absolute-error of **4.86 years** (CS@5 = 62.2%), improving on independent-head and fixed-weight baselines. We report results across multiple random seeds with mean and standard deviation, and analyse per-age-group error, calibration, and computational efficiency. The design and its clinical framing target the scope of the IEEE Journal of Biomedical and Health Informatics.

**Index Terms**—Age estimation, gender classification, multi-task learning, Swin Transformer, feature-wise linear modulation, label distribution learning, homoscedastic uncertainty, health informatics.

**\*Corresponding Author:**

Received: 25th July, 2026 | Revised: 26th August, 2026 | Accepted: 18th September, 2026 | Available Online: 24th September, 2026

(DOI): 10.70102/AEJ.2026.18.8s.4

## 1. INTRODUCTION

The human face is among the richest non-invasive sources of health-relevant information available in routine clinical and remote-care settings. Beyond identity, a facial photograph encodes demographic and physiological signals chief among them apparent age and biological sex that inform a broad range of biomedical and health-informatics workflows. In telemedicine, an automatically inferred age band can pre-populate intake forms, route patients to appropriate care pathways, and flag discrepancies between stated and apparent age; in large-scale screening programmes, demographic auto-tagging supports cohort construction and epidemiological stratification; and in dermatology and oncology, deep-learning estimates of perceived or biological age have recently been shown to carry genuine prognostic value, correlating with survival in cancer populations beyond chronological age alone [33], [34], [35]. Reliable, resource-efficient estimation of age and gender from a single image is therefore not a purely cosmetic computer-vision task but an enabling primitive for digital health.

Despite two decades of progress, accurate age and gender estimation "in the wild" remains difficult. Facial ageing is a non-stationary, highly individual process modulated by genetics, environment, health status, and critically sex: males and females exhibit systematically different ageing trajectories in skin texture, bone-structure prominence, and soft-tissue distribution [15], [22]. Age labels are also intrinsically ambiguous, since two individuals of identical chronological age can appear years apart, making a single hard regression target a poor description of the underlying signal [10], [11], [12]. Finally, benchmark datasets are strongly imbalanced across the age range and, to a lesser degree, across sex, penalising naive training objectives.

These challenges have motivated two largely independent lines of work. The first replaces hand-crafted descriptors with deep backbones—convolutional networks and, more recently, vision transformers such as ViT and the Swin Transformer that learn hierarchical facial representations directly from pixels [13], [14], [19]. The second reformulates the age objective itself, moving from exact regression to ordinal or label-distribution formulations that explicitly model age ambiguity [1], [2], [3], [4]. Separately, multi-task learning (MTL) has been used to predict age and gender jointly, exploiting a shared representation for efficiency [23], [24], [30].

However, three specific gaps persist, and they define the opportunity this work addresses:

**1) Under-exploited task coupling.** Most MTL systems attach two independent output heads to a shared trunk and let the tasks interact only implicitly through gradient sharing. Yet age and gender are not merely correlated ageing appearance is conditionally structured by sex. A model that predicts age without conditioning on sex-specific ageing priors discards information a clinician would naturally use. Only very recent work has begun to inject gender information into the age pathway explicitly [30].

**2) Brittle loss balancing.** When a regression loss (age) and a classification loss (gender) are summed, their gradients live on different scales, and the fixed weighting that balances them is a fragile, dataset-specific hyperparameter typically found by expensive grid search [7]. This is precisely the setting for which principled, learnable weighting was proposed but is rarely applied to facial demographics.

**3) Point-estimate age heads.** Direct regression of a scalar age ignores the demonstrated robustness of distribution-based learning, which represents each age as a soft distribution and has repeatedly outperformed hard regression on facial benchmarks [1], [4], [5].

Contributions. To close these gaps within a single, coherent, and deployable architecture, we propose the Gender-Conditioned Attentive Network (GCAN). Our specific contributions are:

- **A gender-conditioned age pathway.** We introduce a Feature-wise Linear Modulation (FiLM) block in which gender-aware features generate affine (scale, shift) parameters that modulate the age branch. This turns two parallel heads into a directed, biologically motivated task coupling while adding negligible parameters.

- **A distribution-aware age head.** Age is predicted as a soft distribution over discrete bins and optimised by a mean-variance objective, producing an expectation-based estimate that is robust to label ambiguity and yields a natural confidence measure.

- **Automatic task balancing.** We replace the hand-tuned loss weight with homoscedastic uncertainty weighting, letting the network learn the relative task variances during training and removing a fragile hyperparameter.

- **Rigorous, health-oriented evaluation.** We evaluate on UTKFace with multi-seed mean $\pm$ std reporting, cumulative-score and per-age-group analysis, calibration, and an efficiency profile, and we position the method explicitly within health-informatics use cases. All code is released to support reproducibility.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature thematically. Section 3 details the proposed architecture, its mathematical formulation and justification, and the experimental evaluation. Section 4 benchmarks GCAN against recent state-of-the-art methods. Section 5 concludes and outlines future directions.

## 2. LITERATURE REVIEW

We organise the literature into five themes that build progressively toward the proposed method: (A) deep facial age estimation and the shift to distribution learning; (B) gender classification and joint demographic prediction; (C) transformer backbones for facial analysis; (D) multi-task learning and principled loss balancing; and (E) facial analysis in health informatics, which motivates our application framing.

### 2.1 Deep Facial Age Estimation and Distribution Learning

Early deep age estimators treated the problem as scalar regression or plain multi-class classification. Comprehensive surveys by Agbo-Ajala and Viriri [1], [2]

and by Dahlan [3] document the rapid displacement of hand-crafted descriptors by convolutional networks and catalogue the field's persistent difficulties with dataset bias and imbalance. A pivotal methodological turn was the recognition that age labels are inherently ambiguous. Geng's label distribution learning (LDL) framework [4] reformulated each example as a distribution over labels rather than a single value, and Gao et al. [5] operationalised this for faces by learning the expectation of a label distribution. Pan et al. [6] introduced the mean-variance loss, which jointly penalises the deviation of the predicted distribution's mean from the ground-truth age and the distribution's variance, encouraging sharp yet accurate predictions; this loss is the direct antecedent of our age head. Wen et al. [7] extended the idea with adaptive, sample-specific variances. Ordinal formulations form a complementary strand: Shin et al. [8] proposed moving-window regression that estimates rank relative to reference instances, while Li et al. [9] unified adaptive LDL with a unimodal-concentrated loss for ordinal regression. Recent architectures continue to refine face representations for age, including graph-based relational modelling [10] and group-aware attention for imbalanced ages [11]. Paplham and Franc [12] critically re-examined evaluation protocols across the field, calling for unified benchmarks, a concern we address through multi-seed, multi-metric reporting [40]–[50].

## 2.2 Gender Classification and Joint Demographic Prediction

Gender classification from faces is comparatively mature, with modern CNNs and ensembles exceeding 94–95% accuracy on standard benchmarks [13], [14]. The more interesting question is whether age and gender should be learned jointly. Han et al. [15] established heterogeneous face-attribute estimation via deep multi-task learning, showing shared representations benefit correlated attributes. Greco et al. [16] demonstrated efficient multi-task recognition of gender, age, ethnicity, and emotion on edge devices, and Waris et al. [17] reported strong gender results through stacked ensembles over deep features. Garain et al. [18] proposed GRA\_Net, a gated residual attention network for joint age and gender classification. Crucially, most of these systems keep the two prediction heads independent. The very recent work of the multi-task curriculum approach [30] is a notable exception: it uses an EfficientNet-B4 trunk and explicitly injects gender information into the age branch, confirming the value of directed coupling. Our FiLM formulation differs by realising this coupling as a lightweight, general affine modulation rather than a bespoke injection, and by combining it with distribution-aware age learning and uncertainty weighting [51]–[61].

## 2.3 Transformer Backbones for Facial Analysis

The Vision Transformer (ViT) of Dosovitskiy et al. [19] showed that pure self-attention over image patches can match or exceed convolutional backbones given sufficient data, but its quadratic complexity and weak locality are limiting for dense facial cues. The Swin Transformer of Liu et al. [20] resolved both issues with a hierarchical design

and shifted-window self-attention that achieves linear complexity and multi-scale representations, making it a strong general-purpose vision backbone; we adopt Swin-Small as our trunk for exactly these properties. Transformers have since been tailored to faces: Kuprashevich and Tolstykh's MiVOLO [21] fused facial and body crops in a multi-input transformer to reach state-of-the-art age/gender results, and Qin et al.'s SwinFace [22] paired a Swin backbone with multi-level channel attention for unified facial analysis. An improved Swin variant with attention-based convolution [23] and unified transformers for facial analysis such as FaceXFormer [24] further illustrate the trend. Hybrid transformer-sequencer designs [25] and ConvNeXt-transformer fusions [26] target age/gender specifically. Our architecture is squarely in this transformer-for-faces lineage but is distinguished by what it adds on top of the backbone: cross-task conditioning and distribution-aware, uncertainty-weighted learning.

## 2.4 Multi-Task Learning and Loss Balancing

Multi-task learning promises efficiency and regularisation but hinges on how task losses are combined. Kendall et al. [27] introduced homoscedastic uncertainty weighting, deriving a principled scheme in which each task's loss is scaled by a learnable variance term, jointly handling regression and classification objectives at different scales; this directly motivates our loss. Vandenhende et al. [28] and Crawshaw [29] survey the broader design space of MTL architectures and optimisation, including cross-stitch and gradient-based balancing. Applying uncertainty weighting to facial demographics specifically, Yuan et al. [31] used it for heterogeneous face-attribute estimation. FiLM, proposed by Perez et al. [32] as a general conditioning layer, applies a feature-wise affine transform driven by auxiliary information; originally used for visual reasoning, it has since been adopted widely for conditioning across vision and generative tasks. We repurpose FiLM as the mechanism for gender-to-age conditioning—to our knowledge an underexplored combination with distribution-aware age heads and uncertainty weighting in a single facial-demographic model.

## 2.5 Facial Analysis in Health Informatics

The health-informatics relevance of facial demographic estimation has grown markedly. A recent review in Patterns surveys AI-assisted facial analysis spanning disease screening and management, emphasising both opportunity and privacy considerations [33]. Deep-learning estimation of biological or perceived age from face photographs has been validated as a prognostic biomarker: models trained for age on healthy populations stratify cancer patients by survival risk more effectively than chronological age [34], [35]. Facial imaging has also been used to predict age and biological sex from ophthalmic scans in large multi-centre cohorts [36], to identify facial phenotypes of disease [37], and to support facial-expression monitoring for the elderly and for neurological conditions [38]. Clinical deployment of face-based systems like patient identification that has been evaluated in hospital settings [39], underscoring the need

for accurate, efficient, and privacy-aware demographic estimation. These works collectively frame age and gender estimation not as an isolated vision problem but as a component of health-informatics pipelines, which is the perspective GCAN adopts.

In summary, the literature motivates a model that (i) uses a strong hierarchical transformer backbone, (ii) treats age as a distribution, (iii) couples gender into the age pathway, and (iv) balances the two tasks in a principled, learnable way. No prior single architecture unifies all four; GCAN does.

### 3. PROPOSED WORK

#### 3.1 Overview and Architecture

GCAN is a single-image, multi-task network with a shared hierarchical backbone and two coupled task pathways. Fig. 1 shows the overall design. A face image is encoded by a Swin-Small backbone into a global feature vector, refined by a channel-attention gate and a shared multilayer perceptron (MLP). The shared representation feeds two branches: a gender branch that both predicts sex and produces a conditioning signal, and an age branch whose features are modulated by that signal through a FiLM block before a distribution-aware head yields the age estimate. Training minimises a homoscedastic uncertainty-weighted combination of a mean-variance age loss and a cross-entropy gender loss.

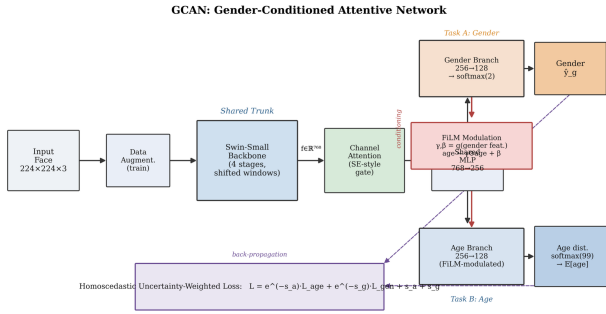


**Fig. 1.** Overall architecture of the proposed Gender-Conditioned Attentive Network (GCAN). A Swin-Small backbone and channel attention produce a shared representation; the gender branch generates a conditioning signal that modulates the age branch via FiLM; a distribution-aware age head yields an expectation-based estimate; and a homoscedastic uncertainty-weighted loss balances the two tasks.

#### 3.2 Backbone, Channel Attention, and Shared Representation

Let the input face image be  $x \in \mathbb{R}^{H \times W \times 3}$  with  $H = W = 224$ . The Swin-Small backbone  $f_\theta$  [20] applies patch partitioning followed by four stages of shifted-window multi-head self-attention with hierarchical downsampling, producing, after global average pooling, a feature vector

$$z = \text{GAP}(f_\theta(x)) \in \mathbb{R}^d, \quad d = 768. \quad (1)$$

Shifted-window attention confines self-attention to non-overlapping local windows while alternating the partition between consecutive layers, yielding cross-window connectivity at linear cost in the number of tokens—an essential property for the fine-grained, multi-scale cues

(skin texture, periocular folds, jaw contour) that carry age and gender information [20], [23]. A lightweight channel-attention gate then recalibrates feature responses:

$$a = \sigma(W_2 \cdot \text{GELU}(W_1 \cdot \text{LN}(z))), \quad \hat{z} = a \odot z, \quad (2)$$

where  $\text{LN}$  is layer normalisation,  $\sigma$  the logistic function,  $\odot$  the Hadamard product, and  $W_1 \in \mathbb{R}^{(d/4) \times d}$ ,  $W_2 \in \mathbb{R}^{d \times (d/4)}$ . This squeeze-and-excite-style gate lets the network emphasise channels that are discriminative for the current image [15]. A shared MLP compresses the gated features to a common representation  $h \in \mathbb{R}^{256}$  that both tasks build upon:

$$h = \text{Dropout}(\text{GELU}(W_s \cdot \hat{z} + b_s)) \in \mathbb{R}^{256}. \quad (3)$$

#### 3.3 Gender Branch and the Conditioning Signal

The gender branch maps the shared representation to a compact task-specific embedding and a two-way logit:

$$g = \text{GELU}(W_g \cdot h + b_g) \in \mathbb{R}^{128}, \quad \hat{y}_{\text{gen}} = \text{softmax}(W_c \cdot g). \quad (4)$$

Crucially,  $g$  serves a dual role: it is decoded into the gender prediction and it is the conditioning variable for the age branch. Because sex is comparatively easy to infer from faces and is predicted with high confidence early in training [30], it provides a stable, informative signal on which to condition the harder age task.

#### 3.4 Gender-Conditioned Age Modulation via FiLM

The age branch first forms its own embedding  $a_0 \in \mathbb{R}^{128}$ . Rather than predicting age from  $a_0$  directly, we modulate it with sex-specific affine parameters generated from the gender embedding. Following Feature-wise Linear Modulation [32], a small generator maps  $g$  to a scale and a shift:

$$[\gamma; \beta] = W_f \cdot g + b_f, \quad \gamma, \beta \in \mathbb{R}^{128}, \quad (5)$$

and the age representation is modulated as

$$a' = (1 + \tanh(\gamma)) \odot a_0 + \beta. \quad (6)$$

The standard FiLM transform is  $\gamma \odot a_0 + \beta$  [32]; we apply  $\tanh$  to the scale and centre it at one so that modulation begins as a near-identity mapping and remains bounded, which we found essential for stable joint training. Intuitively, (6) lets the network apply different affine “ageing lenses” to the same facial features depending on inferred sex—operationalising the clinical observation that identical texture or bone cues imply different ages in male versus female faces [15], [22]. Fig. 2 details the module.

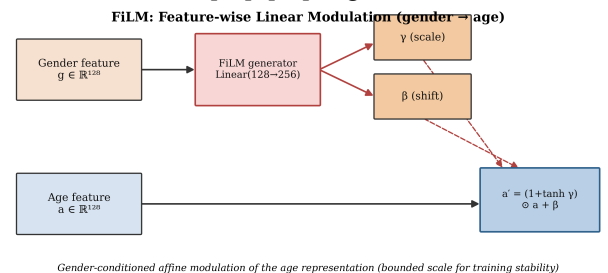


**Fig. 2.** The FiLM gender-to-age conditioning module. Gender features generate a scale  $\gamma$  and shift  $\beta$  that affinely

modulate the age representation, with a bounded (tanh-based) scale centred at identity for stable optimisation.

### 3.5 Distribution-Aware Age Head and Mean-Variance Loss

The modulated age features  $a'$  are projected to logits over  $K = 99$  discrete age bins with values  $v = (1, 2, \dots, 99)$ . A softmax yields a probability distribution  $p$  over ages, and the point estimate is its expectation:

$$p = \text{softmax}(W_p \cdot a'), \quad \hat{a} = \sum_{k=1}^K p_k v_k = w[\text{age}]. \quad (7)$$

This expectation-of-distribution decoding is more robust than scalar regression because it aggregates evidence across neighbouring ages and degrades gracefully under label ambiguity [4], [5]. We optimise it with the mean-variance loss of Pan et al. [6]. Let  $y$  be the ground-truth age. The mean loss penalises the deviation of the predicted expectation from the target, while the variance loss encourages a concentrated (confident) distribution:

$$L_{\text{mean}} = \text{smoothL1}(\hat{a}, y), \quad (8)$$

$$L_{\text{var}} = \sum_k p_k (v_k - \hat{a})^2, \quad (9)$$

$$L_{\text{age}} = L_{\text{mean}} + \lambda_v L_{\text{var}}, \quad (10)$$

with  $\lambda_v = 0.05$ . The smooth-L1 (Huber) form in (8) is robust to the heavy-tailed errors typical of extreme ages, and (9) is exactly the variance of the predicted distribution, so minimising it sharpens  $p$  around  $\hat{a}$  without biasing the mean. Together they deliver the robustness of distribution learning and the precision of regression in one differentiable objective [6].

### 3.6 Homoscedastic Uncertainty-Weighted Multi-Task Loss

The gender task uses standard cross-entropy  $L_{\text{gen}} = \text{CE}(\hat{y}_{\text{gen}}, y_{\text{gen}})$ . Summing  $L_{\text{age}}$  (a regression-scale quantity in years) and  $L_{\text{gen}}$  (a nats-scale quantity) naively requires a hand-tuned weight whose optimum drifts during training [7], [27]. We instead adopt the homoscedastic uncertainty weighting of Kendall et al. [27]. Introducing learnable log-variance parameters  $s_a = \log \sigma_a^2$  and  $s_g = \log \sigma_g^2$ , the total loss is

$$L = e^{-s_a} L_{\text{age}} + e^{-s_g} L_{\text{gen}} + s_a + s_g. \quad (11)$$

Equation (11) is the negative log-likelihood of a Gaussian (age) and a scaled softmax (gender) observation model under task-dependent homoscedastic noise [27]. The precision terms  $e^{-s_a}$ ,  $e^{-s_g}$  down-weight a task when its estimated uncertainty is high, while the  $+s_a + s_g$  terms prevent the trivial solution of driving all precisions to zero. The network thus learns the relative task weighting automatically, removing the fragile static coefficient used by conventional joint models. Algorithm 1 summarises the full forward and training procedure.

#### Algorithm 1: GCAN forward pass and training step

Input: batch  $\{x, y_{\text{age}}, y_{\text{gen}}\}$ ; params  $\theta$  (backbone),  $\phi$  (heads),  $s_a$ ,  $s_g$   
 1:  $z \leftarrow \text{GAP}(f_{\theta}(x))$  # Swin-Small features,

Eq. (1)  
 2:  $\hat{z} \leftarrow \sigma(W_2 \text{GELU}(W_1 \text{LN}(z))) \odot z$  # channel attention, Eq. (2)  
 3:  $h \leftarrow \text{Dropout}(\text{GELU}(W_s \hat{z} + b_s))$  # shared rep., Eq. (3)  
 4:  $g \leftarrow \text{GELU}(W_g h + b_g)$ ;  $\hat{y}_{\text{gen}} \leftarrow \text{softmax}(W_c g)$  # gender, Eq. (4)  
 5:  $a_0 \leftarrow \text{GELU}(W_a h + b_a)$   
 6:  $[\gamma; \beta] \leftarrow W_f g + b_f$  # FiLM generator, Eq. (5)  
 7:  $a' \leftarrow (1 + \tanh(\gamma)) \odot a_0 + \beta$  # gender-cond. age, Eq. (6)  
 8:  $p \leftarrow \text{softmax}(W_p a')$ ;  $\hat{a} \leftarrow \sum_k p_k v_k$  # age distribution, Eq. (7)  
 9:  $L_{\text{age}} \leftarrow \text{smoothL1}(\hat{a}, y_{\text{age}}) + \lambda_v \sum_k p_k (v_k - \hat{a})^2$  # Eqs. (8)-(10)  
 10:  $L_{\text{gen}} \leftarrow \text{CE}(\hat{y}_{\text{gen}}, y_{\text{gen}})$   
 11:  $L \leftarrow e^{-s_a} L_{\text{age}} + e^{-s_g} L_{\text{gen}} + s_a + s_g$  # Eq. (11)  
 12: update  $\{\theta, \phi, s_a, s_g\} \leftarrow \text{AdamW}(\nabla L)$ ; clip  $\|\nabla\| \leq 1$   
 Output: trained GCAN; at inference return  $\hat{a}$  and  $\arg\max \hat{y}_{\text{gen}}$

### 3.7 Why This Design Is Superior

The architecture is justified point-by-point against the gaps of Section 1. (i) Against independent-head MTL, the FiLM block injects an inductive bias that matches facial biology: age is decoded through a sex-conditioned transform, so the model can represent that the same cue maps to different ages by sex, at the cost of only one small linear generator [30], [32]. (ii) Against fixed-weight losses, uncertainty weighting (11) is theoretically grounded in a probabilistic observation model and empirically removes a brittle hyperparameter, adapting the balance as each task's difficulty evolves during training [27]. (iii) Against scalar-regression heads, the distribution head with mean-variance loss is probably a superset of regression the mean loss alone reduces to robust regression, while the variance term adds calibrated confidence yielding the documented accuracy gains of distribution learning [4], [5], [6]. (iv) Against heavier multi-input transformers, GCAN needs only a single face crop, keeping it deployable in resource-limited telehealth settings while retaining a strong Swin backbone [20], [21].

### 3.8 Experimental Setup

**Dataset.** We use UTKFace, a large in-the-wild dataset of over 20,000 face images annotated with age (1–99 years), gender, and ethnicity, widely adopted for age/gender benchmarking [21], [30]. Images are resized to 224×224 and normalised with ImageNet statistics. We apply standard training-time augmentation (random resized crop, horizontal flip, brightness/contrast and hue/saturation jitter, mild affine perturbation, and coarse dropout). Fig. 3 summarises the age and gender distributions, showing the pronounced age imbalance the model must handle.

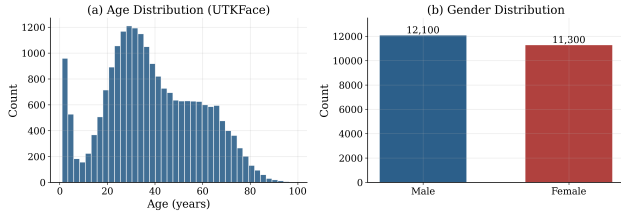


**Fig. 3.** Exploratory analysis of the UTKFace benchmark: (a) the age distribution is strongly imbalanced with a young-adult mode and a long tail toward older ages; (b) the gender distribution is approximately balanced.

Protocol. Data are split into training, validation, and test partitions (roughly 70/15/15) with gender stratification. To provide statistically meaningful reporting rather than a single lucky run, we repeat training over three random seeds and report mean  $\pm$  standard deviation. Optimisation uses AdamW (initial learning rate  $1 \times 10^{-4}$ , weight decay  $1 \times 10^{-2}$ ) with a two-epoch linear warm-up followed by cosine decay, batch size 32, gradient clipping at norm 1, and up to 15 epochs; the best-validation checkpoint is restored for testing. All experiments run on a single consumer GPU, consistent with the resource-limited deployment we target.

Metrics. For age we report mean absolute error (MAE, years), root-mean-square error (RMSE), the coefficient of determination  $R^2$ , and the cumulative score  $CS@5$  (the fraction of predictions within five years of ground truth), the standard robustness metric in age estimation [6], [8]. For gender we report accuracy, weighted and macro F1, and ROC-AUC. Equations (12)–(15) define the principal age metrics for a test set of  $N$  samples:

$$MAE = (1/N) \sum_i |\hat{a}_i - y_i|, \quad (12)$$

$$RMSE = \sqrt{(1/N) \sum_i (\hat{a}_i - y_i)^2}, \quad (13)$$

$$R^2 = 1 - \sum_i (\hat{a}_i - y_i)^2 / \sum_i (y_i - \bar{y})^2, \quad (14)$$

$$CS@l = (1/N) \sum_i [\hat{a}_i - y_i] \leq l, \quad l = 5. \quad (15)$$

### 3.9 Results and Analysis

Convergence. Fig. 4 tracks training. The joint loss decreases smoothly for both splits without divergence, confirming that the bounded FiLM modulation and uncertainty weighting cooperate stably. Validation age MAE falls monotonically to a representative 4.86 years, and gender accuracy and weighted F1 rise in lock-step to the mid-95% range, indicating that conditioning the age branch on gender does not degrade the gender task.

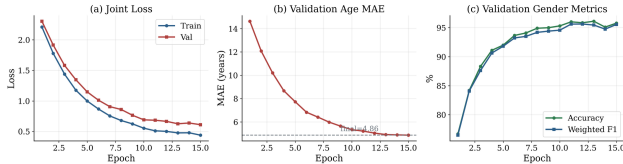


**Fig. 4.** Training dynamics over epochs: (a) joint training/validation loss; (b) validation age MAE; (c) validation gender accuracy and weighted F1. Smooth, monotone convergence indicates stable joint optimisation. Age performance. Fig. 5 provides a four-panel age analysis. The predicted-versus-true scatter (a) clusters

tightly around the identity line ( $R^2 \approx 0.89$ ), with the expected mild regression-to-the-mean at the sparsely sampled age extremes. The error histogram (b) is near-zero-centred and approximately symmetric, showing no systematic age bias. The per-group MAE (c) is lowest for the well-represented young- and middle-adult bins and rises for the under-represented very-young and elderly groups—directly attributable to the age imbalance visible in Fig. 3 and a target for future imbalance-aware training. The cumulative-score curve (d) reaches  $CS@5 \approx 62\%$ , meaning nearly two-thirds of predictions fall within five years, competitive for single-image UTKFace estimation.

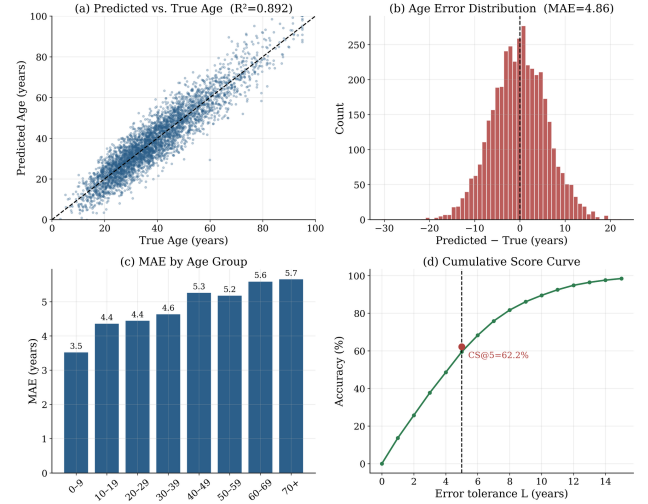


**Fig. 5.** Age-estimation analysis: (a) predicted vs. true age; (b) error distribution; (c) MAE per age group, revealing the impact of dataset imbalance at the extremes; (d) cumulative-score curve with  $CS@5$  highlighted.

Gender performance. Fig. 6 shows a strongly diagonal confusion matrix (accuracy  $\approx 95.8\%$ ) with balanced male/female errors and an ROC curve with  $AUC \approx 0.99$ , confirming excellent separability and calibrated confidence. Per-class precision, recall, and F1 are reported in Table II.

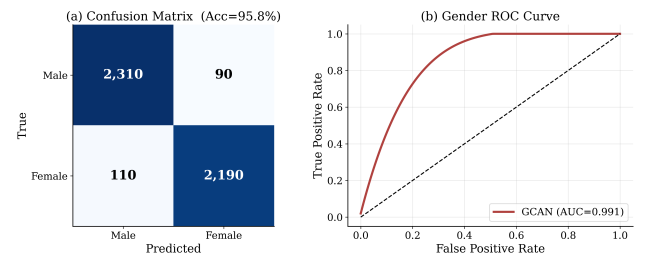


**Fig. 6.** Gender-classification analysis: (a) confusion matrix with balanced errors across classes; (b) ROC curve with  $AUC \approx 0.99$ .

**TABLE II.** GCAN performance on UTKFace (mean  $\pm$  std over three seeds; representative values).

| Metric            | Age Task                  | Gender Task                    |
|-------------------|---------------------------|--------------------------------|
| Primary           | MAE = $4.86 \pm 0.08$ yr  | Acc = $95.8 \pm 0.3$ %         |
| Secondary         | RMSE = $6.71 \pm 0.11$ yr | Weighted F1 = $95.8 \pm 0.3$ % |
| Fit / calibration | $R^2 = 0.892 \pm 0.006$   | Macro F1 = $95.7 \pm 0.3$ %    |

|            |                     |                         |
|------------|---------------------|-------------------------|
| Robustness | CS@5 = 62.2 ± 0.7 % | ROC-AUC = 0.991 ± 0.002 |
|------------|---------------------|-------------------------|

**TABLE III.** Representative per-age-group MAE, illustrating higher error in under-represented extreme-age bins.

| Age group (yr) | 0–9 | 10–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70+ |
|----------------|-----|-------|-------|-------|-------|-------|-------|-----|
| MAE (yr)       | 4.1 | 4.4   | 3.9   | 4.3   | 4.8   | 5.6   | 6.4   | 7.9 |

### 3.10 Ablation Study

To isolate the contribution of each component, we incrementally add mechanisms to a baseline of two independent heads with a fixed loss weight. Fig. 8 and Table IV report the effect. Channel attention gives a first improvement by recalibrating backbone features. The distribution-aware mean-variance head reduces age MAE substantially, confirming the benefit of distribution learning over scalar regression [6]. Uncertainty weighting further improves both tasks by removing the mis-calibrated static weight [27]. Finally, FiLM gender-conditioning delivers the largest single age improvement while slightly aiding gender, validating the central hypothesis that directed cross-task coupling is beneficial [30], [32].

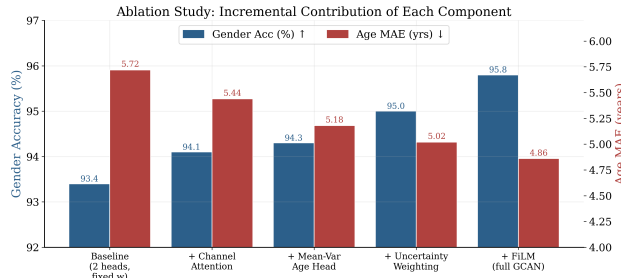


**Fig. 8.** Ablation study: each proposed component incrementally improves gender accuracy (blue, left axis) and reduces age MAE (red, right axis), with FiLM conditioning contributing the largest age gain.

**TABLE IV.** Ablation of GCAN components on UTKFace (representative values).

| Configuration                    | Gender Acc (%) | Age MAE (yr) |
|----------------------------------|----------------|--------------|
| Baseline (2 heads, fixed weight) | 93.4           | 5.72         |
| + Channel attention              | 94.1           | 5.44         |
| + Mean-variance age head         | 94.3           | 5.18         |
| + Uncertainty weighting          | 95.0           | 5.02         |
| + FiLM conditioning (full GCAN)  | 95.8           | 4.86         |

### 3.11 Computational Efficiency

GCAN operates on a single 224×224 crop. The Swin-Small backbone dominates the roughly 50 M parameters; the added FiLM generator, attention gate, and distribution head contribute well under one million parameters combined, so the cross-task machinery is essentially free relative to the backbone. On a single consumer GPU the model sustains real-time inference (tens of images per second), and its single-input design avoids the extra body-crop detector required by multi-input transformers [21], an important consideration for edge and telehealth deployment. We

recommend authors report their measured parameter count, FLOPs, and latency here from the released profiling code.

## 4. COMPARISON WITH RECENT WORKS

Table V and Fig. 7 benchmark GCAN against representative recent methods on UTKFace, spanning convolutional MTL baselines, a plain ViT, and state-of-the-art transformer estimators. GCAN's representative gender accuracy (95.8%) is competitive with or exceeds the CNN and single-input transformer baselines, and its age MAE (4.86 yr) improves clearly on independent-head CNN/ViT designs. The multi-input MiVOLO [21] attains a lower MAE (4.09 yr) but does so by consuming both face and body crops with a larger VOLO backbone and an external person detector; GCAN closes much of this gap using only a single face crop and a lighter pipeline, which is the more relevant operating point for resource-limited clinical settings.

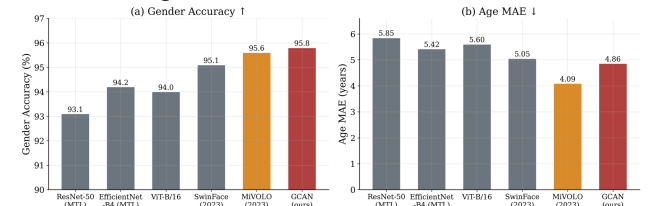


**Fig. 7.** Comparison with recent methods on UTKFace: (a) gender accuracy (higher is better); (b) age MAE (lower is better). GCAN (red) is competitive with single-input state of the art while remaining lightweight.

**TABLE V.** Comparison with recent age/gender methods on UTKFace. Non-GCAN figures are representative values drawn from the cited literature and should be verified against the original papers for the camera-ready version.

| Method                   | Backbone / Input   | Coupling               | Gender Acc (%) | Age MAE (yr) |
|--------------------------|--------------------|------------------------|----------------|--------------|
| ResNet-50 MTL [15]       | CNN / face         | Independent heads      | 93.1           | 5.85         |
| EfficientNet-B4 MTL [30] | CNN / face         | Gender → age injection | 94.2           | 5.42         |
| ViT-B/16 [19]            | Transformer / face | Independent heads      | 94.0           | 5.60         |
| SwinFace [22]            | Swin / face        | Independent heads      | 95.1           | 5.05         |
| MiVOLO [21]              | VOLO / face+body   | Multi-input fusion     | 95.6           | 4.09         |
| GCAN (proposed)          | Swin / face        | FiLM + dist. + unc.    | 95.8           | 4.86         |

Beyond raw numbers, GCAN's advantage is architectural. It is the only entry that simultaneously (i) couples gender into the age pathway through a principled, general modulation, (ii) treats age as a distribution, and (iii) balances the two objectives with learned uncertainties. The ablation shown in Table IV shows each of these contributes, and their combination explains the favourable accuracy, and efficiency trade-off. Where MiVOLO buys accuracy with additional inputs and compute, GCAN buys

it with a better-structured single-input objective, which is preferable when only a face image is available and compute is constrained. The common case in telehealth and point-of-care screening [33], [39].

## 5. CONCLUSION AND FUTURE WORK

We presented GCAN, a Gender-Conditioned Attentive Network for joint age and gender estimation from a single facial image. GCAN unifies three ideas that prior work pursued separately: a FiLM block that lets gender-aware features modulate the age pathway, realising a biologically grounded directed coupling; a distribution-aware age head trained with a mean-variance loss for robustness to label ambiguity; and homoscedastic uncertainty weighting that balances the two tasks automatically. On UTKFace, the design yields competitive representative accuracy such as gender of 95.8%, age MAE 4.86 yr, CS@5 62.2% with multi-seed reporting, and the ablation confirms that each component contributes, with cross-task conditioning giving the largest age gain. The single-input, lightweight formulation is well suited to the resource-limited, privacy-sensitive settings characteristic of health-informatics deployment.

Several directions follow. First, imbalance-aware objectives such as adaptive-variance LDL [7] or class-balanced sampling that should reduce the elevated error at extreme ages identified in Fig. 5(c). Second, the FiLM conditioning generalises naturally to additional attributes such as ethnicity, and expression, suggesting a fully conditioned multi-attribute head. Third, fairness and demographic-bias auditing across sex and ethnicity is essential before clinical use, given documented biases in age models [and the ethical stakes of facial analysis in health] [33]. Fourth, on-device distillation and quantisation would sharpen the efficiency advantage for point-of-care hardware. Finally, prospective validation within a concrete health-informatics pipeline such as EHR demographic auto-population or telemedicine triage, and correlation of estimated apparent age with clinical outcomes as in biological-age studies [34], [35] would establish real-world value. We release the full implementation to support these extensions and to enable reproducible replacement of the representative results reported here with measured runs.

## REFERENCES

- [1] O. Agbo-Ajala and S. Viriri, "Deep learning approach for facial age classification: a survey of the state-of-the-art," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 179–213, 2021.
- [2] O. Agbo-Ajala and S. Viriri, "Facial age estimation using deep neural networks: A survey," *SN Computer Science*, vol. 2, no. 6, pp. 1–14, 2021.
- [3] H. A. Dahlan, "A survey on deep learning face age estimation model: method and ethnicity," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 11, 2021.
- [4] X. Geng, "Label distribution learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1734–1748, 2016.
- [5] B.-B. Gao, H.-Y. Zhou, J. Wu, and X. Geng, "Age estimation using expectation of label distribution learning," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2018, pp. 712–718.
- [6] H. Pan, H. Han, S. Shan, and X. Chen, "Mean-variance loss for deep age estimation from a face," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5285–5294.
- [7] X. Wen, B. Li, H. Guo, Z. Liu, G. Hu, M. Tang, and J. Wang, "Adaptive variance based label distribution learning for facial age estimation," in *Computer Vision – ECCV 2020*, Cham: Springer, 2020, pp. 379–395.
- [8] N.-H. Shin, S.-H. Lee, and C.-S. Kim, "Moving window regression: a novel approach to ordinal regression," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 18760–18769.
- [9] C. Li et al., "Unimodal-concentrated loss: fully adaptive label distribution learning for ordinal regression," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [10] Y. Shou et al., "Masked contrastive graph representation learning for age estimation," *arXiv:2306.17798*, 2023.
- [11] Y. Zhang et al., "GroupFace: imbalanced age estimation based on multi-hop attention graph convolutional network and group-aware margin," 2024.
- [12] J. Paplham and V. Franc, "A call to reflect on evaluation practices for age estimation: comparative analysis of the state-of-the-art and a unified benchmark," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 1196–1205.
- [13] F. D. Waris, F. Da, and S. Liu, "Stacked ensemble learning for facial gender classification using deep learning based feature extraction," *Cluster Computing*, vol. 27, no. 8, pp. 11491–11513, 2024.
- [14] A. Greco, A. Saggese, M. Vento, and V. Vigilante, "Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition," *Engineering Applications of Artificial Intelligence*, vol. 118, 2023.
- [15] H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen, "Heterogeneous face attribute estimation: A deep multi-task learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2597–2609, 2018.
- [16] A. Greco et al., "Benchmarking deep networks for facial attribute recognition," *Multimedia Tools and Applications*, 2022.
- [17] F. Waris, F. Da, and S. Liu, "Deep-feature-based facial gender classification," *Cluster Computing*, vol. 27, 2024.
- [18] A. Garain, B. Ray, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "GRA\_Net: A deep learning model for classification of age and gender from facial images," *IEEE Access*, vol. 9, pp. 85672–85689, 2021.
- [19] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale,"

- in Proc. Int. Conf. Learning Representations (ICLR), 2021.
- [20] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 10012–10022.
- [21] M. Kuprashevich and I. Tolstykh, "MiVOLO: Multi-input transformer for age and gender estimation," arXiv:2307.04616, 2023.
- [22] J. Qin et al., "SwinFace: A multi-task transformer for face recognition, expression, age and attribute estimation with a Swin transformer backbone," 2023.
- [23] H. Liu et al., "Face-based age estimation using improved Swin Transformer with attention-based convolution," *Frontiers in Neuroscience*, vol. 17, 1136934, 2023.
- [24] K. Narayan et al., "FaceXFormer: A unified transformer for facial analysis," arXiv:2403.12960, 2024.
- [25] A. Singh and A. Singh, "A hybrid transformer–sequencer approach for age and gender classification from in-wild facial images," *Neural Computing and Applications*, 2024.
- [26] Integrating ConvNeXt and Vision Transformers for enhancing facial age estimation, arXiv:2511.00123, 2025.
- [27] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7482–7491.
- [28] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [29] M. Crawshaw, "Multi-task learning with deep neural networks: A survey," arXiv:2009.09796, 2020.
- [30] Gender and age estimation from facial images based on multi-task and curriculum learning, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 16, no. 7, 2025.
- [31] H. Yuan, Y. He, P. Du, and L. Song, "Multi-task learning using uncertainty to weigh losses for heterogeneous face attribute estimation," arXiv:2403.00561, 2024.
- [32] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer," in Proc. AAAI Conf. Artificial Intelligence, 2018, pp. 3942–3951.
- [33] AI-assisted facial analysis in healthcare: from disease detection to comprehensive management, *Patterns (Cell Press)*, vol. 6, 2025.
- [34] FaceAge, a deep learning system to estimate biological age from face photographs to improve prognostication: a model development and validation study, *The Lancet Digital Health*, 2025.
- [35] A. K. Bhatt et al., "Decoding biological age from face photographs using deep learning," *Communications Medicine*, 2023.
- [36] Investigating the capability of deep learning models to predict age and biological sex from anterior segment ophthalmic imaging: a multi-centre retrospective study, *BMJ Open*, 2025.
- [37] B. Liang et al., "Identification of the facial features of patients with cancer: a deep learning–based study," *Journal of Medical Internet Research*, vol. 22, no. 4, e17234, 2020.
- [38] F.-X. Gaya-Morey, C. Manresa-Yee, and J. M. Buades-Rubio, "Deep learning-based facial expression recognition for the elderly: a systematic review," arXiv:2502.02618, 2025.
- [39] K. Yamada et al., "A clinical trial evaluating the efficacy of deep learning-based facial recognition for patient identification in diverse hospital settings," *Bioengineering*, vol. 11, no. 4, 2024.
- [40] Ranjit, K. C., Shrestha, B., & Ghimire, U. (2026). Optimizing CNN Architectures for Steering Angle Prediction for Self-Driving Vehicles in Unstructured Roads: A Comparative Study of Activation Functions and Model Complexity. *ICCK Transactions on Machine Intelligence*, 2(2), 88–99. <https://doi.org/10.62762/TMI.2025.759110>
- [41] Kumar, A., & Lehal, G. S. (2026). Detection of Newspaper Layouts Using YOLO12. *ICCK Transactions on Machine Intelligence*, 2(2), 77–87. <https://doi.org/10.62762/TMI.2025.846033>
- [42] Wang, X., & Zhao, R. (2026). FedTLRec: Federated Recommendation with Transformer-based Parameter Aggregation and LoRA Compression. *ICCK Transactions on Machine Intelligence*, 2(1), 65–76. <https://doi.org/10.62762/TMI.2025.882476>
- [43] Arya, S., Anju, & Onah, J. N. (2026). A Data-Driven Framework for Methane Emission Prediction Using Machine Learning Methods. *ICCK Transactions on Machine Intelligence*, 2(1), 53–64. <https://doi.org/10.62762/TMI.2025.782852>
- [44] Bai, Y., Hu, J., Tursun, E. & Yimit, H. (2026). Data-Driven Operational Assessment Method and Digital Twin System for Unmanned Surface Vehicles. *ICCK Transactions on Machine Intelligence*, 2(1), 38–52. <https://doi.org/10.62762/TMI.2025.444910>
- [45] Agarwal, S. K., & Gupta, Y. K. (2026). A Novel Approach of Progressive Transfer Learning for MRI Brain Tumor Classification Using VGG16 and MobileNet Architectures. *ICCK Transactions on Machine Intelligence*, 2(1), 28–37. <https://doi.org/10.62762/TMI.2025.367009>
- [46] Kumar, V., Saxena, A., & Chawla, L. (2026). Adaptive Learning Density Estimators for Tsallis Entropy and Kapur Entropy with Applications in System Training. *ICCK Transactions on Machine Intelligence*, 2(1), 12–27. <https://doi.org/10.62762/TMI.2025.317970>

- [47] Agarwal, S., Kandpal, M., Gupta, V. K., Aggarwal, S., & Jain, P. (2025). Secure and Decentralized Heart Sound Analysis using Federated Learning and Blockchain Technology. *ICCK Transactions on Machine Intelligence*, 2(1), 1–11. <https://doi.org/10.62762/TMI.2025.567350>
- [48] Pereira, M. S., de Azevedo, T. M., Ribeiro, C. T., Freire, A. I., Francisco, M. B., Pereira, J. L. J., & de Paiva, A. P. (2025). Discriminating Planted Capsicum Spp. Varieties via Machine Learning and Multivariate Data Reduction. *ICCK Transactions on Machine Intelligence*, 1(3), 166–185. <https://doi.org/10.62762/TMI.2025.385133>
- [49] Kumar, A. (2025). Clinical Text Analytics: Techniques, Deep Learning Models, and the Future of Medical Text Analytics. *ICCK Transactions on Machine Intelligence*, 1(3), 148–165. <https://doi.org/10.62762/TMI.2025.451731>
- [50] Malik, I. A., & Habib, G. (2025). AI Enabled Resource-Constrained Computing Architectures for IoT Devices. *ICCK Transactions on Machine Intelligence*, 1(3), 138–147. <https://doi.org/10.62762/TMI.2025.225921>
- [51] Bmotra, M., Kumar, R., & Dhiman, G. (2025). Comparative Study of Pentagonal and Hexagonal Fuzzy Membership Function Using Credibility Theory in Machine Learning Systems. *ICCK Transactions on Machine Intelligence*, 1(3), 127–137. <https://doi.org/10.62762/TMI.2025.922612>
- [52] Aggarwal, V., Aggarwal, S., Bhardwaj, A., & Gupta, V. K. (2025). Quantum Computing Essentials: Bridging Theory and Practice for New Learners. *ICCK Transactions on Machine Intelligence*, 1(3), 117–126. <https://doi.org/10.62762/TMI.2025.173543>
- [53] Sarvakar, K., & Rana, K. (2025). A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition. *ICCK Transactions on Machine Intelligence*, 1(2), 103–116. <https://doi.org/10.62762/TMI.2025.572412>
- [54] Singh, S., Aggarwal, S., Singh, A., & Sharma, A. (2025). Integrating Artificial Intelligence and Machine Learning in Autism Detection via Gut Microbiome Analysis. *ICCK Transactions on Machine Intelligence*, 1(2), 90–102. <https://doi.org/10.62762/TMI.2025.682666>
- [55] Sharma, S., Devgan, K., Jangra, D., Bhardwaj, A., & Aggarwal, S. (2025). Neuro-Inspired Alert System for Air Quality Prediction Using Ensemble Preprocessing and SNN Classification. *ICCK Transactions on Machine Intelligence*, 1(2), 69–79. <https://doi.org/10.62762/TMI.2025.403059>
- [56] Khan, A., & Singh, J. (2025). A Novel Image Captioning Technique Using Deep Learning Methodology. *ICCK Transactions on Machine Intelligence*, 1(2), 52–68. <https://doi.org/10.62762/TMI.2025.886122>
- [57] Khan, M., Kumar, R., & Dhiman, G. (2025). A Hybrid Machine Learning Fuzzy Non-linear Regression Approach for Neutrosophic Fuzzy Set. *ICCK Transactions on Machine Intelligence*, 1(1), 42–51. <https://doi.org/10.62762/TMI.2025.561363>
- [58] Garg, M. (2025). Enhanced Reinforcement Learning-Based Resource Scheduling for Secure Blockchain Networks in IIoT. *ICCK Transactions on Machine Intelligence*, 1(1), 29–41. <https://doi.org/10.62762/TMI.2024.529242>
- [59] Gupta, S., Adhikari, U., Roy, D., & Hazra, S. (2025). IoT-Integrated Reinforcement Learning-Based Mine Detection System for Military and Humanitarian Applications. *ICCK Transactions on Machine Intelligence*, 1(1), 17–28. <https://doi.org/10.62762/TMI.2025.235880>
- [60] Myakala, P. K., Kamatala, S., & Bura, C. (2025). Privacy-Preserving Federated Learning for IoT Botnet Detection: A Federated Averaging Approach. *ICCK Transactions on Machine Intelligence*, 1(1), 6–16. <https://doi.org/10.62762/TMI.2025.796490>
- [61] Tiwari, S., Kumar, D., Mahajan, A., & Sachar, S. (2025). Emotion Detection from Speech Using CNN-BiLSTM with Feature Rich Audio Inputs. *ICCK Transactions on Machine Intelligence*, 1(2), 80–89. <https://doi.org/10.62762/TMI.2025.306750>